Identifying Cell Type-Specific Chemokine Correlates with Hierarchical Signal Extraction from Single-Cell Transcriptomes*

Sherry Chao

Department of Biomedical Informatics, Harvard University
Boston, MA, United States
E-mail: schao@g.harvard.edu

Michael P. Brenner

School of Engineering and Applied Sciences, Harvard University
Cambridge, MA, United States
E-mail: brenner@seas.harvard.edu

Nir Hacohen

Harvard Medical School, Harvard University
Boston, MA, United States
E-mail: nhacohen@mgh.harvard.edu

Biological data is inherently heterogeneous and high-dimensional. Single-cell sequencing of transcripts in a tissue sample generates data for thousands of cells, each of which is characterized by upwards of tens of thousands of genes. How to identify the subsets of cells and genes that are associated with a label of interest remains an open question. In this paper, we integrate a signal-extractive neural network architecture with axiomatic feature attribution to classify tissue samples based on single-cell gene expression profiles. This approach is not only interpretable but also robust to noise, requiring just 5% of genes and 23% of cells in an *in silico* tissue sample to encode signal in order to distinguish signal from noise with greater than 70% accuracy. We demonstrate its applicability in two real-world settings for discovering cell type-specific chemokine correlates: predicting response to immune checkpoint inhibitors in multiple tissue types and classifying DNA mismatch repair status in colorectal cancer. Our approach not only significantly outperforms traditional machine learning classifiers but also presents actionable biological hypotheses of chemokine-mediated tumor immunogenicity.

Keywords: Interpretable machine learning; Translational cancer research; Single-cell RNA-sequencing; Chemokines.

1. Introduction

The advent of technologies to sequence tissue samples at single-cell resolution has ushered in a new era of biological learning.[1] The ability to characterize intercellular variability with high granularity not only furthers our understanding of complex living systems but also necessitates

*This work is supported by the Simons Foundation.

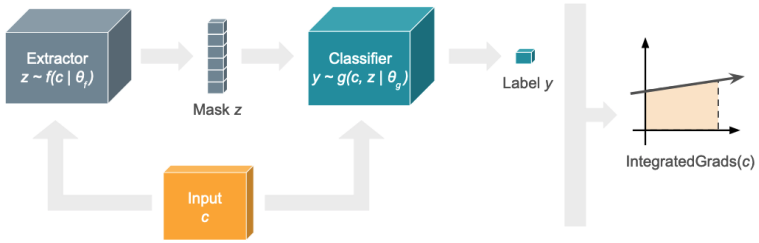

Fig. 1. **Model overview.** The signal extractor takes in a single sample c, comprised of the gene expression profiles of multiple cells, and generates per-cell masks z. The classifier takes in a masked version of the input to generate a label prediction y. Finally, the trained model is interrogated using Integrated Gradients with respect to the original input of gene expression profiles.

novel computational approaches to effectively utilize these additional layers of information. This necessity is amplified in cancer, where tumor cells manipulate their microenvironment by harnessing the plasticity of the cells around them. For instance, while the immune system normally protects the host, tumor cells can cajole certain immune cells into allowing the tumor to not only proliferate but also metastasize.[2] Because tumor cells co-opt immune cell plasticity, heterogeneity in the tumor is greater than that in normal tissue.[3] Understanding cellular heterogeneity via high-dimensional single-cell RNA-sequencing (scRNA-seq) and its implications for the interplay between malignant cells and immune cells is key to treating complicated diseases such as cancer.[4,5]

From a machine learning perspective, the task of classifying tissue samples based on single-cell transcriptomics is a weakly-supervised learning problem. Each data point (tissue sample) consists of a bag of observations (cells), where each observation is characterized by a set of features (genes). The label of interest is assigned to the tissue sample instead of the cell. Therefore, every cell may not contain evidence of the label of interest, but when pooled, observations of many cells together reveal compelling correlations with the label. Standard pooling mechanisms, such as mean-pooling and max-pooling, obfuscate biologically-relevant information. For example, a cell can be characterized by a number of discrete cell types, each of which is capable of existing along a continuum of cell states. Given cells have different types and states, how does the model learn to identify signal? In biology, in particular, the problem is a hierarchical one, since once the cells (and their associated types/states) of interest are identified, we need to find a way to extract the cell-specific genes that reveal associations with the label of interest and ultimately produce accurate classifications.

To address these questions, we present a novel framework for extracting hierarchical information from a neural network classifier that optimizes for both strong classification performance and accurate signal extraction (Figure 1). The signal extraction task is directly incorporated into both the design decisions around model architecture and the model training decisions for downstream feature attribution. The model, a Single-Cell Immuno-Oncology Neural Network (SCIONN), enables accurate classification and signal extraction from structured biological data by identifying a subset of cells of importance and the subset of genes of importance from the aforementioned subset of cells. We employ a learning scheme to quantify the importance of each cell and each cell-specific gene, drawing inspiration from rationale

generation and gradient integration, respectively.[6,7] The benefit of this approach lies in being agnostic to cell type and invariant to the ordering of observations. Unlike with standard machine learning classifiers, such as logistic regression, we do not need to know the cell types *a priori* or prearrange the order in which the cells are fed into the model.

When applied to binary classification on a simulated scRNA-seq dataset and two real-world problems in oncology, the SCIONN classifier outperforms not only logistic regression but also more complex neural network classifiers, such as variants of recurrent neural networks. On the simulated dataset, the SCIONN classifier is far better able to identify the signal-carrying cells and genes. When varying the signal-to-noise, this approach is robust to the number of genes and cells, requiring only 23% of cells in a positive-labeled tissue sample to encode the signal to be highly discriminative, even if just 5% of genes encode the signal. Moreover, performance persists not only when signal is present in multiple cell types but also when signal is present in mutually exclusive gene sets from each of the signal-carrying cell types. We demonstrate the applicability of this approach to two real-world tasks: (1) predicting response to immune checkpoint inhibitors in multiple tissue types and (2) classifying the DNA mismatch repair status of colorectal cancers for determining prognosis and treatment. Using multiple publicly-available datasets for each task, we show compelling evidence for cells and genes associated with these labels, thereby facilitating mechanistically-relevant, clinically-actionable learning.

2. Method

2.1. *Objective*

Let dataset $\mathcal{D}$ consist of N sequenced tissue samples, each of which is represented by the gene expression profiles of a set of N_i cells $c_i \in \mathbb{R}^{N_i \times G}$ and a binary label y_i. Each cell $c_{ij} \in \mathbb{R}^G, j \in \{1, \ldots, N_i\}$, is represented by a vector of gene expression values for G genes. Our objective is to maximize the likelihood function $p(Y \mid C, \theta)$ with respect to parameters θ, where the predictive distribution $p(y \mid c)$ is parameterized by a neural network.

2.2. *Model*

2.2.1. *SCIONN*

SCIONN is a deep neural network consisting of a signal extractor and a classifier (Figure 1). The signal extractor comprises two convolutional layers followed by two recurrent layers and a final fully-connected layer. The classifier comprises two two-dimensional convolutional layers (kernel size (1, 1) and filter size (256, 128)) followed by three fully-connected layers. The input to the signal extractor is the gene expression profiles of cells from a given tissue sample. The input to the classifier is a masked version of the input to the signal extractor, as determined by the signal extractor output z, where z is an N_i-dimensional vector and each element is either zero or one. The signal extractor output masks, or zeroes out, the gene expression profiles of a subset of input cells and retains the rest, as given by $(zv^T) \circ c_i$, where v is a G-dimensional vector of ones. The signal extractor finds the cells of interest, and the classifier finds the genes of interest from the cells of interest.

2.2.2. *Benchmarks*

We compare the classification performance of SCIONN to traditional machine learning classifiers, which do not perform any masking, namely logistic regression (LogReg) and three variants of two-layer recurrent neural networks - a vanilla recurrent neural network (RNN), a long short term memory network (LSTM), and a gated recurrent unit network (GRU). For logistic regression, the gene expression profiles of C cells are concatenated along the existing axis to form the model input; for all other models, the gene expression profiles are shaped into a $C \times G$ matrix. Unless otherwise specified, the size of all hidden layers is 64.

2.3. *Inputs*

Weakly-supervised learning on scRNA-seq data suffers from two problems: (1) there is no guaranteed signal from single cells, and (2) the number of labeled samples n is typically small ($n < 100$), while the number of features k is large ($k > 100{,}000$, e.g., 1,000 cells $\times$ 100 genes). To address these problems, we process the input data by creating pseudo-samples of randomly sampled sets of 100 cells from the same tissue sample, where each set of 100 cells is a pseudo-sample. Cells are sampled without replacement unless there are less than 100 cells from a particular tissue sample. This sampling structure assumes that the proportion of different cell types in the pseudo-sample mimics the proportion of different cell types in the tissue sample in expectation. With pseudo-samples, we increase the effective diversity of the dataset – akin to data augmentation – in order to help the model learn with greater robustness and efficiency.

2.4. *Training*

We randomly split the dataset into train (80%), validation (10%), and test (10%) sets and standardize the features of each set to the train set. The splits are made at the patient level to prevent information leakage, and model inputs are constructed at the tissue sample level, as there can be multiple tissue samples for the same patient. Models are trained end-to-end for 200 epochs with a batch size of 200, learning rate of 0.0001, and 50% dropout at every intermediate fully-connected layer. We employ an Adam optimizer to minimize the binary cross entropy loss.[8] We reduce the learning rate by a factor of 0.1 if the loss does not improve after ten consecutive epochs. In order to enable efficient learning during SCIONN training, we sample from a gumbel-softmax distribution parameterized by the signal extractor outputs and a temperature of 3.0.[9,10] We regularize the number of cells selected by the signal extractor with a lambda of 0.0001 to enforce sparsity. Beginning at training epoch 50, the temperature is reduced by 0.01 and lambda is increased by 0.1 if the loss does not improve after ten consecutive epochs. The weights of the model with the lowest binary cross entropy loss on the validation set are saved and subsequently evaluated on the held-out test set. We repeat this process for 50 random train/validation/test splits for all classifiers.

2.5. *Attribution*

Achieving good classification with SCIONN ensures the model has found a correlation structure of the data with respect to the label. To probe this correlation structure and identify

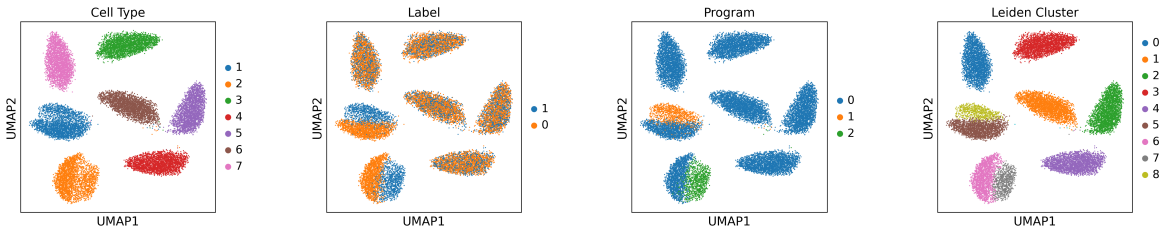

Fig. 2. **Simulated data overview.** The simulated data comprises seven cell types, of which two encode signal. Samples containing cells belonging to cluster 7 or 8 (i.e., having non-zero gene program assignment) are assigned a positive label.

important cell type-specific genes, we compute the attribution scores using integrated gradients, a method that is both feature sensitive and implementation invariant.[7] The integrated gradients method integrates over the gradients with respect to the label from a baseline to the input. For our purposes, the baseline across all genes is zero, standardized to the train set.

2.5.1. *Baseline Training*

To ensure the baseline is not biased for or against either label in binary classification, during training, half of the inputs the model "sees" are a baseline of zeroes standardized to the train set. Half of these baseline examples are given the label 0, and the remaining half are given the label 1. This strategy ensures that the model achieves a classification AUC of 0.50 on the baseline examples alone, thereby rendering the model unbiased to the chosen baseline.[11]

2.6. *Experiments*

We apply our framework under three settings:

(1) Binary classification with simulated scRNA-seq data where positive-labeled samples are determined by two cell types expressing either the same or different signal-specific genes.
(2) Immune checkpoint inhibitor response prediction from the chemokine expression profiles of single cells from three publicly-available scRNA-seq datasets.
(3) DNA mismatch repair status classification from the chemokine expression profiles of single cells from three publicly-available scRNA-seq datasets.

3. Results

3.1. *Simulated Data*

We sought to determine whether SCIONN could accurately identify the features contributing to the label of interest by simulating a scRNA-seq dataset of 20,000 cells and 464 genes.[12] Under the splat generative process, single-cell gene expression is modeled according to a Gamma-Poisson model, accounting for factors that influence gene expression in real-world scRNA-seq datasets: highly-expressed outliers, library size, mean-variance trends, and dropout. The splat model outperformed five models on three real-world datasets on seven evaluation metrics.

Table 1. **Simulated data results.** Performance mean ± 1SD on held-out test set. '*', '+', and '⋆' signify SCIONN outperforms LogReg, RNN, and CNN, respectively, at the 0.05 level (Wilcoxon one-sided signed-rank test).

Gene Set	Model	Loss	AUC	Number of Correct Genes as Fraction of Top 120 Genes	Number of Correct Genes as Fraction of Genes $> \mu + 2\sigma$
Same	LogReg	0.2470 ± 0.1415	0.9813 ± 0.0720	0.8137 ± 0.1251	0.9257 ± 0.1523
	RNN	0.1331 ± 0.1891	0.9863 ± 0.0399	0.2210 ± 0.0700	0.2518 ± 0.0850
	CNN	0.1516 ± 0.0132	1.0000 ± 0.0000	0.5820 ± 0.0666	0.6302 ± 0.0876
	SCIONN	**0.1858 ± 0.0325** *	**1.0000 ± 0.0000** * +	**0.8160 ± 0.0215** + ⋆	**0.9067 ± 0.0303** + ⋆
Different	LogReg	0.3812 ± 0.1560	0.9436 ± 0.1272	0.6812 ± 0.1477	0.8121 ± 0.1965
	RNN	0.2822 ± 0.3654	0.9539 ± 0.0862	0.1983 ± 0.0590	0.2193 ± 0.0687
	CNN	0.1710 ± 0.0209	1.0000 ± 0.0000	0.6103 ± 0.0407	0.7065 ± 0.0871
	SCIONN	**0.2261 ± 0.0382** *	**1.0000 ± 0.0000** * +	**0.7740 ± 0.0321** * + ⋆	**0.8976 ± 0.0340** * + ⋆

In our simulated dataset, each cell belongs to one of seven cell types. While all samples contain cells from every cell type, only two cell types determine the label of interest (cell types 1 and 2, Figure 2). Namely, the presence of cells from clusters 7 and 8 (subsets of cell types 1 and 2, respectively) in a given sample signifies that the sample is a positive-labeled example; otherwise, the absence of cells from these clusters signifies that the sample is a negative-labeled example. Our objective is to evaluate whether SCIONN is able to (1) find the correct cells (i.e., clusters 7 and 8) via the signal extractor, and (2) find the correct genes that distinguish clusters 7 and 8, thereby enabling accurate binary classification.

We tested two scenarios of signal-specific genes (programs), of which there are 60 from cluster 7 and 60 from cluster 8. In one scenario, the signal-specific genes of clusters 7 and 8 are the same. The differentially expressed genes from these two clusters are partially cell type-specific (characteristic of cell types 1 and 2, respectively) and partially signal-specific, where clusters 7 and 8 share the same signal-specific genes. In the second scenario, the signal-specific genes of clusters 7 and 8 are different. The differentially expressed genes from these clusters are partially cell type-specific (characteristic of cell types 1 and 2, respectively) and partially signal-specific, but the signal-specific genes differ between clusters 7 and 8 (programs 1 and 2, Figure 2). In the second scenario, we are introducing the added challenge of identifying not only the subsets of cells of interest but also their mutually exclusive genes of interest. This setup is important because oftentimes different cell types have different transcriptional responses to the same stimuli, and we want to correctly capture the knowledge that these different responses are associated with the same label of interest.

We trained SCIONN on this simulated dataset under both gene set scenarios and compared its performance against a logistic regression classifier and an RNN classifier (Table 1). On the basis of binary cross entropy loss, SCIONN significantly outperformed the logistic regression classifier while exhibiting performance on par with that of the RNN classifier (same gene set: 0.1858 vs. 0.2470 and 0.1331, respectively; different gene set: 0.2261 vs 0.3812 and 0.2822, respectively). This result suggests that choice of architecture is important; here, the architecture needs to be agnostic to the order of cells, which is true for SCIONN and the RNN classifier but not for the logistic regression classifier. On the basis of AUC, SCIONN significantly outperformed both the logistic regression classifier and the RNN classifier, achieving

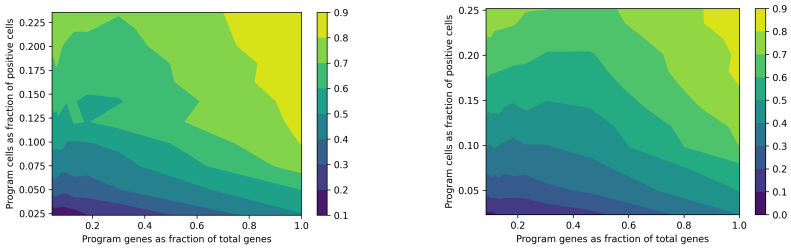

Fig. 3. **Sensitivity analysis.** Signal extraction accuracy as a function of the number of signal-specific genes (as a fraction of total genes) and the number of signal-specific cells (as a fraction of total cells from positive-labeled samples).

perfect average AUC (Same gene set: 1.000 vs 0.9813 and 0.9863, respectively; different gene set: 1.000 vs 0.9436 and 0.9539, respectively).

After establishing that SCIONN yields strong classification performance, we next turned to the task of assessing whether the model learned the 60 signal-specific genes from cluster 7 and the 60 signal-specific genes from cluster 8. Using the integrated gradients method to compute the attribution scores, we subsequently average over the attribution scores for a given cell type and gene and rank the cell type-gene tuples across the positive-labeled samples by descending average attribution. If the model finds the correct cells (cells from clusters 7 and 8 from cell types 1 and 2, respectively) and the correct genes from these correct cells, then we expect these to be 120 signal-specific cell type-gene tuples populating the top attributions. Indeed, we find that SCIONN significantly outperforms the RNN classifier when the between-cluster ground truth genes are the same set of genes, identifying 81.60% of the correct cell type-gene tuples in the top 120 cell type-gene tuples (vs. 22.10% for the RNN classifier). SCIONN performs on par with the logistic regression classifier, which identifies 81.37% of the correct cell type-gene tuples, because logistic regression only cares about the genes of interest. Since the signal-specific genes are the same in clusters 7 and 8, it is able to attribute its classification to those genes irrespective of cell type.

In contrast, when the between-cluster ground truth genes are different sets of genes, SCIONN significantly outperforms both the logistic regression classifier and the RNN classifier (68.12% and 19.83%, respectively), identifying 77.40% of the correct cell type-gene tuples in the top 120 cell type-genes pairs (Table 1). Logistic regression breaks down under this scenario because it does not know how to differentiate between cell types, whereas SCIONN does. Thus, under our more complex gene set scenario, SCIONN retains its ability to successfully identify not only the cells of interest but also the cell-specific genes of interest. Under normal conditions, we do not know *a priori* which are the cells and genes of interest. Therefore, we also assess the extent to which the signal-specific cell type-gene tuples populate the top attributions, defined as the set of cell type-gene tuples whose attribution exceeds two standard deviations above the mean attribution across all tuples. Again we find the same pattern as before. SCIONN performs on par with the logistic regression classifier and significantly outperforms the RNN classifier when the between-cluster ground truth genes are the same set of genes (90.67% vs 92.57% and 25.18%, respectively). SCIONN significantly outperforms both the logistic regression classifier and the RNN classifier when the between-cluster ground truth

Table 2. **PD-1 response prediction results.** Performance mean ± 1SD on held-out test set. '∗', '⋄', '⋆', and '∘' signify SCIONN outperforms LogReg, RNN, LSTM, and GRU, respectively, at 0.05 level (Wilcoxon one-sided signed-rank test).

Model	Loss	AUC
LogReg	0.6322 ± 0.0652	0.7119 ± 0.1116
RNN	0.6484 ± 0.0643	0.6443 ± 0.1340
LSTM	0.6362 ± 0.1063	0.6889 ± 0.1662
GRU	0.6292 ± 0.0900	0.7028 ± 0.1387
SCIONN	**0.6144 ± 0.0884** ∗⋄⋆∘	**0.7583 ± 0.1588** ∗⋄⋆∘

Table 3. **MMR classification results.** Performance mean ± 1SD on held-out test set. '∗', '⋄', '⋆', and '∘' signify SCIONN outperforms LogReg, RNN, LSTM, and GRU, respectively, at 0.05 level (Wilcoxon one-sided signed-rank test).

Model	Loss	AUC
LogReg	0.6849 ± 0.1173	0.6040 ± 0.2477
RNN	0.6878 ± 0.0204	0.5408 ± 0.0782
LSTM	0.6250 ± 0.0795	0.7573 ± 0.1292
GRU	0.6406 ± 0.0732	0.7023 ± 0.1258
SCIONN	**0.5320 ± 0.0679** ∗⋄⋆∘	**0.8701 ± 0.0954** ∗⋄⋆∘

genes are different sets of genes (89.76% vs 81.21% and 21.93%, respectively).

Finally, we run a sensitivity analysis on the fraction of signal-specific cells and genes needed to achieve certain levels of detection. Under the scenario where the between-cluster ground truth genes are the same set of genes, we find that if approximately 20% of positive-labeled cells are signal-specific cells, then SCIONN is able to accurately identify at least 70% of signal-specific cell type-gene tuples when less than 30% of genes are signal-specific genes (Figure 3, left panel). Under the scenario where the between-cluster ground truth genes are different sets of genes, we find that if approximately 25% of positive-labeled cells are signal-specific cells, then SCIONN is able to accurately identify at least 70% of signal-specific cell type-gene tuples when less than 20% of genes are signal-specific genes (Figure 3, right panel). These results are encouraging, as there is often a far greater fraction of signal-specific cells than genes in any given sample given the sheer number of profiled genes.

3.2. *Real-World Data*

3.2.1. *Background*

Chemokines are chemotactic cytokines that mediate cell migration and positioning in both tissue and lymph nodes, particularly with respect to immune cells.[13,14] Once a chemokine receptor is activated by its cognate chemokine ligand, the chemokine receptor-expressing cell migrates up the associated chemokine concentration gradient.[13,15] In the context of cancer, chemokines can promote pro-tumorigenic activity, such as tumor proliferation, anti-apoptosis, and metastasis.[15,16] Furthermore, chemokines can mediate immune evasion by influencing T cell sequestration in stroma or by cultivating a microenvironment of high-density non-activating immune cells, leading to inefficient T cell search of malignant cells. The importance of chemokines to tumorigenicity warrants further study of these molecules, their role in the tumor-immune microenvironment, and their role in response to immunotherapy.

3.2.2. *PD-1 Response Prediction*

To study the role of chemokines in response to immune checkpoint inhibitors, we integrated the chemokine expression profiles from three scRNA-seq datasets.[17–19] We corrected for batch effects associated with individual studies for the 64 chemokines and chemokines receptors.[20] We trained on immune checkpoint inhibitor response prediction based on chemokine expres-

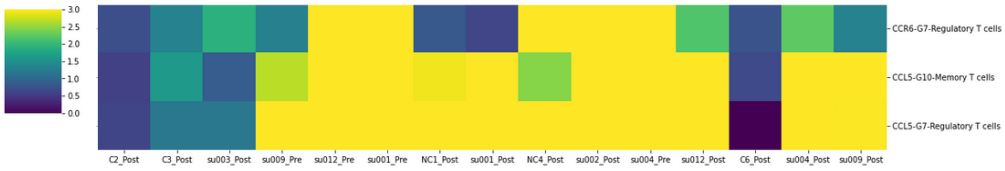

Fig. 4. PD-1 response prediction top ranked attribution scorers by (gene, cell type).

sion profiles from all available timepoints. The binary response label was defined based on RECIST; a responder (R) has a complete or partial response, and a non-responder (NR) has stable or progressive disease.[21] We compared five model architectures - a logistic regression classifier, vanilla RNN classifier, LSTM classifier, GRU classifier, and SCIONN. On the basis of both binary cross entropy loss and AUC, SCIONN significantly outperformed the four other classifiers (mean loss: 0.6144 vs. 0.6322, 0.6484, 0.6362, and 0.6292, respectively; mean AUC: 0.7583 vs. 0.7119, 0.6443, 0.6889, and 0.7028, respectively) (Table 2).

Using the trained SCIONN model, we subsequently explored the attribution scores of all cell type-gene tuples. T cell subtypes for GSE123813 and GSE144469 were mapped to the subtypes annotated by Ref 18, while all other immune subtype labels were retained as-is.[22] We excluded epithelial cells for consistency across the datasets, as some datasets sorted for CD45+ cells prior to performing single-cell sequencing. Based on the attribution scores, we found the majority of attribution was assigned to chemokines, consistent with the view that chemokines exhibit greater variability of expression compared to their receptor counterparts.[23] Furthermore, expression of the chemokine CCL5 in the memory T cell (Tmem) and regulatory T cell (Treg) compartments consistently scored high on attribution for immune checkpoint inhibitor responders (Figure 4). This finding is consistent with recent findings from *in vivo* models demonstrating that intratumorally-administered CCL5 enhanced cytotoxic lymphocytes and the anti-tumor activity of anti-PD-L1.[24] The association with Tmem and Treg subpopulations in particular suggest that the immune system is not only mounting a recurrent response (via activated Tmem) but also a new response (via activated Treg) to the ever-evolving tumor. In addition, we identified a novel feature associated with response, namely CCR6 expression in Tregs, which warrants further study. Given CCR6 is a receptor and receptor expression is typically stable, this particular finding in favor of CCR6 in the Treg subpopulation suggests active migration of these cells towards the tumor.

3.2.3. *Mismatch Repair Classification*

Next, we studied the implications of DNA mismatch repair status on chemokine activity. DNA mismatch repair deficiency (MMRd) is a notable cancer phenotype due to its association with response to immune checkpoint inhibitors, but the mechanism of response is unclear.[5] We trained SCIONN on DNA mismatch repair status classification using the chemokine expression profiles of a 60-patient colorectal cancer cell atlas dataset.[3] We compared SCIONN's performance to that of four other classifiers - a logistic regression classifier, vanilla RNN classifier, LSTM classifier, GRU classifier. On the basis of both binary cross entropy loss and AUC, SCIONN significantly outperformed the four other classifiers (mean loss: 0.5320 vs.

Table 4. **MMR classification results on independent datasets**. Performance mean ± 1SD. '∗', '⋄', '⋆', and '○' signify SCIONN outperforms LogReg, RNN, LSTM, and GRU, respectively, at the 0.05 level (Wilcoxon one-sided signed-rank test).

Independent Dataset	Model	Loss	AUC
GSE146771_colon10x	LogReg	1.3730 ± 0.6203	0.4934 ± 0.1323
	RNN	0.7004 ± 0.0555	0.4993 ± 0.1495
	LSTM	0.6425 ± 0.1082	0.6194 ± 0.2475
	GRU	0.5878 ± 0.1222	0.6419 ± 0.1914
	SCIONN	**0.3920 ± 0.1021** ∗⋄⋆○	**0.8691 ± 0.2201** ∗⋄⋆○
GSE146771_colonSS2	LogReg	0.7878 ± 0.1469	0.5328 ± 0.0842
	RNN	0.6967 ± 0.0313	0.5302 ± 0.0859
	LSTM	0.6623 ± 0.0811	0.7197 ± 0.1864
	GRU	0.6431 ± 0.1001	0.6848 ± 0.1603
	SCIONN	**0.5030 ± 0.0897** ∗⋄⋆○	**0.9041 ± 0.0843** ∗⋄⋆○
GSE132465_colon10x	LogReg	0.8509 ± 0.1503	0.5469 ± 0.0541
	RNN	0.6974 ± 0.0346	0.5074 ± 0.0676
	LSTM	0.7574 ± 0.1474	0.6333 ± 0.0821
	GRU	0.6915 ± 0.0823	0.5950 ± 0.0964
	SCIONN	**0.6711 ± 0.0960** ∗⋄⋆	**0.7275 ± 0.0321** ∗⋄⋆○

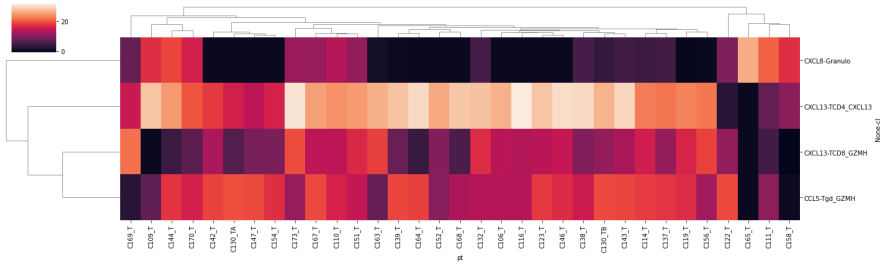

Fig. 5. MMR classification top ranked attribution scorers by (gene, cell type).

0.6849, 0.6878, 0.6250, and 0.6406, respectively; mean AUC: 0.8701 vs. 0.6040, 0.5408, 0.7573, and 0.7023, respectively) (Table 3). We validated the models on two independent colorectal cancer scRNA-seq datasets (GSE146771 and GSE132465).[25,26] SCIONN's classification performance persisted on these held-out independent datasets, significantly outperforming the other four models. When we evaluated the trained SCIONN classifier on these datasets, SCIONN achieved an average binary cross entropy loss range of 0.3920-0.6711 compared to 0.5878-1.3730 across the other four models and an average AUC range of 0.7275-0.9041 compared to 0.4934-0.7197 across the other four models (Table 4).

Using the trained SCIONN model, we subsequently explored the attribution scores of all cell type-gene tuples. Immune and stromal subtypes for GSE146771 and GSE132465 were mapped to the subtypes annotated by Ref 3 for the colorectal cancer cell atlas.[22] We excluded epithelial cells during training for consistency across the datasets. Based on the attribution scores, we found that the majority of attribution for MMRd was attributable to T cells, which is consistent with the view that MMRd induces greater immunogenicity given its propensity for mutated neoantigens.[4] CXCL13 in activated CD4+ and CD8+ T cell subpopulations

were assigned the highest attribution, which persisted even in the held-out external datasets, consistent with findings from Ref 3 (Figure 5). Furthermore, CCL5 in GZMH^{hi} CD8+ T and $\gamma\delta$T cell populations were also assigned high attributions, particularly in the colorectal cancer cell atlas dataset. Interestingly, CCL5 in T cell subpopulations was also highlighted in immune checkpoint inhibitor response prediction, suggesting that CCL5 is associated with an immunogenic tumor microenvironment, regardless of whether immunogenicity arose from immunotherapy or MMRd. Finally, subsets of tissue samples exhibited high attribution for CXCL8 in granulocytes, which has yet to be explored.

4. Discussion

In this paper, we introduce a framework for identifying cell type-specific chemokine correlates of response to immune checkpoint inhibitors in multiple tissues and DNA mismatch repair status in colorectal cancers. By employing a deep neural network that couples signal extraction with label classification, we are able to successfully identify signals derived from cell type-gene tuples that would otherwise have been overlooked under poorer-performing neural network architectures. Notably, we find this approach to be superior for PD-1 immune checkpoint inhibitor response prediction and DNA mismatch repair status classification based on single-cell chemokine expression profiles. The resulting attribution scores not only validate known biology (e.g., CXCL13 expression in activated T cells in MMRd tumors) but also present new biological hypotheses (e.g., CCR6 expression in Tregs in PD-1 inhibitor responders and CXCL8 expression in granulocytes in MMRd tumors) and suggest a shared chemotactic mechanism of immunogenicity in both PD-1 inhibitor responders and MMRd tumors (CCL5 expression in T cells). We foresee this framework extending to other classification problems where the correlation structure is learned from hierarchically-structured data and are currently working to extend it to spatially-resolved single-cell transcriptomics.

Software and Data

The scRNA-seq datasets were downloaded from GEO with accession numbers: GSE178341,[3] GSE123814,[17] GSE120575,[18] GSE144469,[19] GSE146771,[25] and GSE132465.[26] The code for reproducing all results is publicly available: `https://www.github.com/chsher/SCIONN`.

References

1. B. Lim, Y. Lin and N. Navin, Advancing cancer research and medicine with single-cell genomics, *Cancer cell* **37**, 456 (2020).
2. A. K. Palucka and L. M. Coussens, The basis of oncoimmunology, *Cell* **164**, 1233 (2016).
3. K. Pelka, M. Hofree, J. H. Chen, S. Sarkizova, J. D. Pirl, V. Jorgji *et al.*, Spatially organized multicellular immune hubs in human colorectal cancer, *Cell* (2021).
4. I. Vitale, E. Shema, S. Loi and L. Galluzzi, Intratumoral heterogeneity in cancer progression and response to immunotherapy, *Nature medicine* , 1 (2021).
5. M. Binnewies, E. W. Roberts, K. Kersten, V. Chan, D. F. Fearon, M. Merad, L. M. Coussens, D. I. Gabrilovich, S. Ostrand-Rosenberg, C. C. Hedrick *et al.*, Understanding the tumor immune microenvironment (time) for effective therapy, *Nature medicine* **24**, 541 (2018).

6. T. Lei, R. Barzilay and T. Jaakkola, Rationalizing neural predictions, *arXiv preprint arXiv:1606.04155* (2016).
7. M. Sundararajan, A. Taly and Q. Yan, Axiomatic attribution for deep networks, in *International Conference on Machine Learning*, 2017.
8. D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).
9. E. Jang, S. Gu and B. Poole, Categorical reparameterization with gumbel-softmax, *arXiv preprint arXiv:1611.01144* (2016).
10. C. J. Maddison, A. Mnih and Y. W. Teh, The concrete distribution: A continuous relaxation of discrete random variables, *arXiv preprint arXiv:1611.00712* (2016).
11. K. McCloskey, A. Taly, F. Monti, M. P. Brenner and L. J. Colwell, Using attribution to decode binding mechanism in neural network models for chemistry, *PNAS* **116**, 11624 (2019).
12. L. Zappia, B. Phipson and A. Oshlack, Splatter: simulation of single-cell rna sequencing data, *Genome biology* **18**, 1 (2017).
13. J. W. Griffith, C. L. Sokol and A. D. Luster, Chemokines and chemokine receptors: positioning cells for host defense and immunity, *Annual review of immunology* **32**, 659 (2014).
14. K. Schumann, T. Lämmermann, M. Bruckner, D. F. Legler, J. Polleux, J. P. Spatz, G. Schuler, R. Förster, M. B. Lutz, L. Sorokin *et al.*, Immobilized chemokine fields and soluble chemokine gradients cooperatively shape migration patterns of dendritic cells, *Immunity* **32**, 703 (2010).
15. M. F. Krummel, F. Bartumeus and A. Gérard, T cell migration, search strategies and mechanisms, *Nature Reviews Immunology* **16**, p. 193 (2016).
16. M. T. Chow and A. D. Luster, Chemokines in cancer, *Cancer immunology research* **2**, 1125 (2014).
17. K. E. Yost, A. T. Satpathy, D. K. Wells, Y. Qi, C. Wang, R. Kageyama, K. L. McNamara, J. M. Granja, K. Y. Sarin, R. A. Brown *et al.*, Clonal replacement of tumor-specific t cells following pd-1 blockade, *Nature medicine* **25**, 1251 (2019).
18. M. Sade-Feldman, K. Yizhak, S. L. Bjorgaard, J. P. Ray, C. G. de Boer, R. W. Jenkins, D. J. Lieb, J. H. Chen, D. T. Frederick, M. Barzily-Rokni *et al.*, Defining t cell states associated with response to checkpoint immunotherapy in melanoma, *Cell* **175**, 998 (2018).
19. A. M. Luoma, S. Suo, H. L. Williams, T. Sharova, K. Sullivan, M. Manos *et al.*, Molecular pathways of colon inflammation induced by cancer immunotherapy, *Cell* **182**, 655 (2020).
20. W. E. Johnson, C. Li and A. Rabinovic, Adjusting batch effects in microarray expression data using empirical bayes methods, *Biostatistics* **8**, 118 (2007).
21. E. A. Eisenhauer, P. Therasse, J. Bogaerts, L. H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney *et al.*, New response evaluation criteria in solid tumours: revised recist guideline (version 1.1), *European journal of cancer* **45**, 228 (2009).
22. F. A. Wolf, P. Angerer and F. J. Theis, Scanpy: large-scale single-cell gene expression data analysis, *Genome biology* **19**, 1 (2018).
23. S. Minina, M. Reichman-Fried and E. Raz, Control of receptor internalization, signaling level, and precise arrival at the target in guided cell migration, *Current biology* **17**, 1164 (2007).
24. Y. W. Lim, G. L. Coles, S. K. Sandhu, D. S. Johnson, A. S. Adler and E. L. Stone, Single-cell transcriptomics reveals the effect of pd-l1/tgf-β blockade on the tumor microenvironment, *BMC biology* **19**, 1 (2021).
25. L. Zhang, Z. Li, K. M. Skrzypczynska, Q. Fang, W. Zhang, S. A. O'Brien, Y. He, L. Wang, Q. Zhang, A. Kim *et al.*, Single-cell analyses inform mechanisms of myeloid-targeted therapies in colon cancer, *Cell* **181**, 442 (2020).
26. H.-O. Lee, Y. Hong, H. E. Etlioglu, Y. B. Cho, V. Pomella, B. Van den Bosch, J. Vanhecke, S. Verbandt, H. Hong, J.-W. Min *et al.*, Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer, *Nature Genetics* **52**, 594 (2020).