



**HAL**  
open science

# Detecting gene regulatory network rewiring as changes in marker regulatory associations using only single-cell gene expression data

Lisa Chabrier, Anton Crombach, Sergio Peignier, Christophe Rigotti

## ► To cite this version:

Lisa Chabrier, Anton Crombach, Sergio Peignier, Christophe Rigotti. Detecting gene regulatory network rewiring as changes in marker regulatory associations using only single-cell gene expression data. IEEE 12th Data Mining in Biomedical Informatics and Healthcare Workshop (DMBIH'24), 2024, Abu Dhabi, United Arab Emirates. pp.10. hal-04808464

**HAL Id: hal-04808464**

**<https://inria.hal.science/hal-04808464v1>**

Submitted on 28 Nov 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Detecting gene regulatory network rewiring as changes in marker regulatory associations using only single-cell gene expression data

Lisa Chabrier<sup>1, 2</sup>, Anton Crombach<sup>1, 2</sup>, Sergio Peignier<sup>3</sup>, and Christophe Rigotti<sup>1, 2</sup>

<sup>1</sup>Inria, Centre de Lyon, 69603 Villeurbanne, France.

<sup>2</sup>INSA Lyon, CNRS, Universite Claude Bernard Lyon 1, LIRIS, UMR5205, 69621 Villeurbanne, France.

<sup>3</sup>INSA Lyon, INRAE, BF2I, UMR0203, 69621, Villeurbanne, France.

## Abstract

Understanding changes in gene regulatory networks, known as rewiring, is a challenge. Methods addressing this problem require multi-omics input data, which is costly to generate for model organisms, and simply not available for non-model organisms. We propose a new approach to detect rewiring events between cell types, that uses only single-cell RNA sequencing data. The approach consists of training machine learning models to predict the expression of target genes from the expression of transcription factors. Then a local (i.e., per cell), model-agnostic explainability method, SHAP, determines the importance of each transcription factor for each target gene. We consider that these transcription factor importances capture the gene network state of a cell. A new space is created describing cells in terms of regulatory network associations between genes. We compute in this space the statistically different associations between cell types, resulting in a set of marker regulatory associations. Several visualizations help the selection of putative rewiring events between GRN states in different cell types. The proposed workflow is applied to two datasets, demonstrating the potential for detecting rewiring. The implementation of our workflow follows the conventions of the scverse initiative and allows easy integration with Scanpy and other packages for single-cell analysis.

**Keywords:** Explainable ML, SHAP, gene regulatory network, network rewiring, differential network analysis.

## 1 Introduction

Cellular processes such as differentiation and environmental adaptation are governed to an important extent by regulatory interactions between genes. At the molecular level such interactions are initiated by Transcription Factors (TFs), which are genes coding for proteins that bind to the DNA and influence the expression of other genes. These interactions can be represented by Gene Regulatory Networks (GRNs). The static, “complete” GRN associated with an entire organism is not fully ac-

tive in every cell. Different TFs will alternatively regulate genes, for instance depending on the organ a cell is located in. These changes in GRNs are called rewiring of GRNs and represent the modification of the topology of the GRN in different conditions and cells. Studies suggest that rewiring events have strong phenotypic impacts [1,2]. The activity of the TFs is cell-type-specific [3], depending on binding partners, chromatin accessibility, and cofactors [4]. Moreover, these rewiring events are influencing a large number of biological processes and are driving phenotypic variations [2,5,6]. Rewiring events are involved in cell fate decisions [7], adaptation to the environment [8], reaction to stress [9], and cell division [10]. The increasing number of published studies that mention regulatory relations helps understand the relevance of rewiring events. For example, TRRUST is a database collecting all the regulatory interactions for humans [11]. In this database, genes are regulated by different TFs depending on the studies, suggesting that indeed rewiring events are very common. Therefore, understanding the changes in GRNs is a central challenge in biology.

The detection of rewiring could help the understanding of the differences between cell types, cell states, disease states, or even treatment effects. Detecting the rewiring events is a key challenge in the understanding of cellular processes. The gold standard *experimental* approach to detect such events is chromatin immunoprecipitation followed by sequencing (ChIP-seq) that measures TF-binding events. When a spiking normalization method is applied, this method is able to detect global changes in TF activity [12]. The main disadvantage, however, is that it provides information for only one TF at a time and it does not scale well to large numbers of TFs. Experiments take up to several days and are costly. In addition, errors in the experimental process occur and experiments may have to be conducted several times. Due to the limitations of this technique, people are turning toward computational methods. Classically, these *in silico* approaches detect rewiring based on the analysis of bulk RNA-Seq, such as [13,14]. Recent high-throughput technologies are opening new perspectives for computational approaches,

using data from techniques such as single-cell omics.

In order to computationally capture rewiring events, different approaches have been proposed that can be grouped into two categories. The methods of the first category are based on the accurate inference of GRNs associated with different populations of cells. The difference between these GRNs is then computed to provide information about the regulatory interactions that changed. This is called differential analysis. This kind of analysis heavily relies on the performance of GRN inference methods. Such methods are well-established and are able to infer satisfactory GRNs, though it remains an active research field and inference methods keep improving. The methods of the second category are fully dedicated to the computation of rewiring and are explicitly computing GRNs for each condition. They often involve complex data science techniques and rely on multi-omics input data [15]. These data offer a lot of insight into biological processes, but they are not available for every experiment and every species. In particular, it is important to distinguish model species—human, mouse, and fruit fly—for which one can have access to multi-omics experiments and data, and pioneer, non-model species, for which such access is not available.

In order to provide a simple solution to as many studies as possible, we chose to focus on single-cell RNA sequencing (scRNA-seq) datasets. These data provide in-depth information about the heterogeneity of gene expression of the studied cell populations. The method we propose is available for studies with only scRNA-seq data. In addition to an scRNA-seq dataset, three parameters are needed. First, a list of TFs for the species of interest. It can be obtained from another species if not available, as TFs are highly conserved through evolution. Second, a clustering of the cell population in cell types, which is a standard functionality offered by most scRNA-seq analysis software. Third, our method will only analyze the rewiring events for a list of genes of interest. Genes of interest can be a small number of genes for which the user suspects an important role or a wider list of genes for an unbiased study. The parameters are easily obtained and allow the method to be useful in various contexts.

Our work is inspired by one of the standard approaches to infer GRNs with only scRNA-seq: GRNBoost2 [16]. In this algorithm, machine learning models are trained to predict the expression of target genes, using the expression of TFs as features. Then, a global feature ranking method (i.e. explainability of machine learning) is applied to each machine learning model to rank the most important regulators of each gene. The link formed by the regulator and the target gene is called a regulatory association, because GRNBoost2 only detects co-expression and an analysis of the binding sites around genes is necessary to conclude that an interaction between the TF and the target gene is possible. Because the feature ranking method is global, a single GRN is inferred for the whole cell population.

We propose to apply a local feature ranking method to similar machine learning models and rank the regulators of a gene, this time for each cell separately. This allows

us to compute a *state of the GRN* for a single cell; it is a vector containing importance scores, describing the *activity* of the regulatory associations. This *state of the GRN* representation of cells is considered as a new space in which each cell is represented by a vector of activities of regulatory associations. We call this space the *regulatory association space*.

Several local explanation methods are available, of which the most popular ones are SHAP values [17] and LIME [18]. In this work, we choose the SHAP values framework, because of its strong mathematical foundation. Nevertheless, SHAP values computation has an exponential complexity in the number of features of the models, therefore computing the exact SHAP values is not considered feasible. Approximation methods have been proposed for the computation of the SHAP values to remove this constraint. And when we approximate SHAP values, we compute confidence intervals and use those to safely select the top- $k$  most important features, including ties. This selection is based on the observation that for each explanation corresponding to each cell, only a subset of TFs (features) will have a high associated SHAP value. For our experiments,  $k$  was set to 10 based on the observation that most of the target genes appear to have less than 10 regulators [19].

The method as outlined above and described in detail in this work is implemented as a simple yet powerful tool to detect key regulators and rewiring events in a GRN, given only a scRNA-seq dataset, a list of TFs, a list of genes of interest, and a clustering of the cell population. We apply our method to publicly available reference datasets with different cell types to identify key regulators (TFs). Next, we provide qualitative analyses of the results obtained for each dataset. We show that the method is able to detect different regulators of the same marker genes, in different clusters, and that the patterns revealed are coherent with the levels of expression of the marker genes studied.

In this article, we first present an overview of the related work. We will then detail the proposed method and the experimental results. A discussion about the results and a conclusion will end this article.

## 2 Related work

Two possible workflows to detect rewiring events are (1) the computation of differential networks from previously inferred GRNs or (2) the direct detection of rewiring.

This first category of methods is called differential network analysis. It is important to note that differential analysis of GRNs can provide more general information than only rewiring detection. Nevertheless, it is possible to use the output of such methods and apply post-processing steps to extract the putative rewiring events. These tools are part of the large group of methods proposing downstream analysis for GRN inference methods. Despite the progress made in recent years, inferring accurate GRNs is still a challenging task, and downstream analysis methods are highly dependent on the quality of the GRN inferred. In differential network analysis, a very

Table I: Input data required by the existing methods.

Method	scRNA-seq	prior network	scATAC-Seq	ChIP-Seq
ANANSE	✓	Human protein atlas	✓	✓
sc-compreg	✓(two populations)		✓(two populations)	
RNCMI	✓	PPI network		
CEFCON	✓	protein-protein interactions		
TopicNet	✓			✓
<b>Our rewiring detection method</b>	✓			

intuitive approach is the direct comparison between two networks, resulting in the intersection operation on the two sets of links of each network. Such an approach has been used successfully in several studies to search for specific regulators (e.g. [20–23]). It has also been used to find regulatory links conserved in evolution across species [24]. Another approach consists of summarizing the topology of the network into a more robust “core” network [25]. This method is used successfully in studies to identify rewiring events in humans GRN [26]. Due to the nature of the inference methods and the noisiness of the initial data, ensemble operation over the list of regulatory links is considered as not sufficiently robust [15] despite the good results described in the different studies. . The second category regroups the methods specifically designed to detect rewiring events. The ANANSE [27] method, for example, computes in parallel several differential networks and computes similarities between the differential networks. The similarities are “usual changes”, condensed into potential rewiring events. The methods use a Gaussian graphical model, D-trace loss is used to compute the differential networks, and a tree-structured group Lasso penalty to identify the common hub nodes in the differential networks. In the topicNet method, latent Dirichlet allocation is used to compute a rewiring score between networks, by summarizing networks to the *main features* and comparing the reduced networks. They apply the method to predict survival in various cancers. The method sc-compreg [28] uses a likelihood ratio statistic to compute the differential regulatory regulations, between two conditions with linked cell types in the two datasets. finally, the method RNCMI [29] key regulators in cell fate decision are detected, based on multi-omics integration. A background network is used and edges are deleted to transform it into a cell-type-specific network. A diffusion analysis is performed to detect differentially regulated genes in the specific networks.

The input data required by these methods are composed of several omic data and prior knowledge of the GRNs. CEFCON [23] and RNCMI [29] are two recently published methods that take as input data scRNA-seq matrices and previous network topology. Most of the methods require additional input data type, including scATAC-seq for sc-compReg [28], and ANANSE [27], and ChIP-seq for ANANSE and TopicNet [30].

However, due to different input data requirements (see Table I) and the need to apply post-processing to differential analysis methods, it is rare to see articles presenting results about the potential rewiring events in the regulatory network. This paper proposes a method using

scRNA-seq and requiring only a list of TFs, a list of genes of interest, and groups of cells for which we wish to detect rewiring. This is applicable to many studies. Our method is implemented as a module compatible with Scanpy [31], a popular data analysis tool in the scRNA-seq community.

### 3 Proposed Method

In this section, we introduce the prerequisites for the computation of SHAP values and then detail the method that we designed to detect GRN rewiring, named *rewiring detection method*.

#### 3.1 SHAP prerequisite

SHAP is an additive feature attribution function and is the unique score that respects fairness axioms, as described in [17]. Each feature is attributed a contribution to the value predicted by the model, either a positive or a negative one. The sum of the values attributed to all features (here the TFs) is the difference between the average output of the model, and the predicted value for this instance (here a cell), as represented in Fig. 1. This property is called additivity.

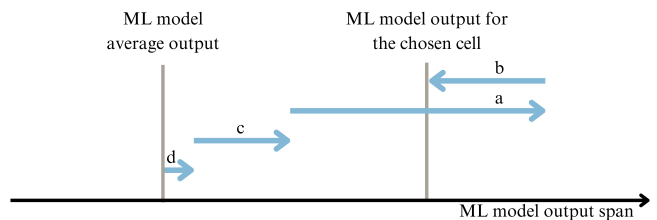


Figure 1: Visual representation of the additivity of the SHAP values for a prediction model using four features (a, b, c, d).

Computing the exact SHAP values requires a number of operations exponential in the number of features [32], thus recent approximation methods are preferred. Therefore we propose to compute the confidence interval of the approximations of the SHAP values and to select the top 10 with ties.

In our approximation framework, SHAP values are computed by sampling the conditional contribution of the feature and taking the mean. This is the approximation framework presented in [33], which shows that these values are normally distributed around the SHAP

value. Using the central limit theorem, it is possible to compute confidence intervals.

We compute the mean and standard deviation of the sampled conditional contributions and then compute a confidence interval. As many confidence intervals are compared, we want to ensure an overall confidence of  $\gamma = 0.95$ . To achieve this, we use the Bonferroni correction adapted to confidence intervals, as described in [34]. The corrected confidence level is  $\gamma' = 1 - \frac{1-\gamma}{|F|}$ . For a cell, the features that are not selected are set to a value of zero in the association matrix.

We present the selection for an example with four features, and the selection of a top-two in Fig. 2. Two values are considered tied if their absolute values cannot be ordered when considering the confidence intervals. Note that we select the most important *absolute* SHAP values, as negative SHAP values are equally important as positive ones.

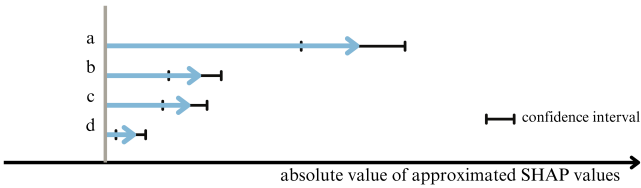


Figure 2: Selection of the top-k, plus ties, most important *absolute-valued* features. Two values are ties if their ranking is ambiguous, when inspecting confidence intervals. Here, when selecting the top-2, features a, b, and c will be selected, because b and c are ties; d is not selected.

### 3.2 Proposed method for the detection of rewiring

The inputs of the proposed *rewiring detection method* are a scRNA-seq dataset, a list of transcription factors (the TFs), and a list of genes of interest. It also requires the different groups of cells for which we want to study the rewiring. These groups of cells can be interpreted as cell types.

The proposed *rewiring detection method* relies on the computation of the association space (step one), that represents the cells by local activity of the cell’s GRN. Then (step two), this association space is used to detect the regulatory associations that are different in a cell type compared to the rest. The detailed *rewiring detection method* is explained below.

#### 3.2.1 Step one

A scRNA-seq dataset is used to train several machine-learning models. We use random forests, with 100 estimators, a squared error criterion, and a sampling of  $\sqrt{n}$  features among  $n$  features available for each split. Each model learns the relationship between the expression of a gene of interest and the expression of the transcription factors (the features), as presented in Fig. 3. Three hypotheses are necessary:

1. It is possible to model the expression of a gene as a function of the expression of the transcription factors.
2. ML models are able to learn this function.
3. With explainability methods (XML), it is possible to access the importance of the features for each cell

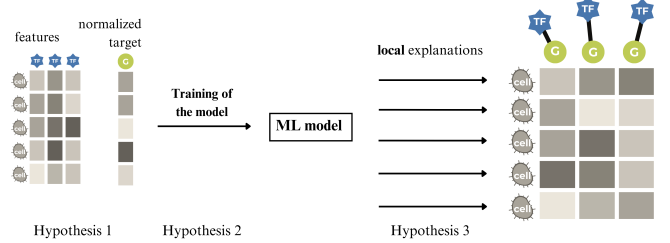


Figure 3: Proposed method for 1 target gene.

The workflow presented in Fig. 3 starts by training an ML model to predict the expression of the gene of interest from the expression of TFs. The SHAP top-k approximation and selection method is applied to the model to compute the feature (TF) importance in the prediction of the expression of the gene of interest for each cell. The resulting matrix contains for all cells a score for the regulatory association between the TFs and the gene of interest. The SHAP values, called association scores, are grouped in an association matrix (see Fig. 4). The standardization (z-scoring) of the expression of the gene of interest before the training of the model ensures that the domain of the output of the model is standard. Because the SHAP values have the same unit and domain as the domain of output of the model (see Fig. 1), with a standardized domain, we can compare the SHAP values computed for different models. In other words, we can compare the association scores computed for different genes of interest. As only the most important features are relevant for this application, a selection of the top 10 SHAP values is made using the absolute SHAP values, but the values reported in the matrix are the original values (negative and positive). Indeed, the sign of the SHAP values indicates the direction of the influence of the feature on the output of the model and we want to keep this information. The scores of the not-selected regulatory associations are set to zero, which is the neutral value.

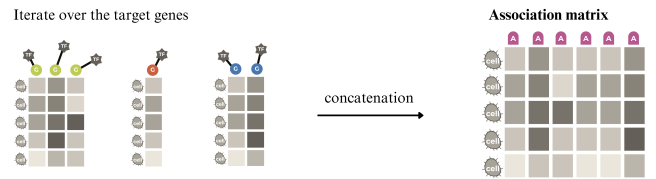


Figure 4: Iterate the process over all the marker genes to compute the association matrix.

### 3.2.2 Step two

We use the association space to compute the most differentially activated regulatory associations for each cell type. We consider two distributions of association score (SHAP values): the association scores in the cell type and the association scores out of the cell type. We use a student’s t-test to test the null hypothesis that the two (independent) distributions have the same mean, compared to the hypothesis that the mean of the cell type distribution is higher. Other hypotheses could be used to detect regulatory associations that are active in all cell types except the one of interest.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad (1)$$

with  $\bar{x}_i$  the mean of the distribution,  $s_p$  the pooled standard deviation, and  $n_i$  the number of elements in the distribution. The t-score is used to describe how different the regulatory association is, in and out of the cell type. The regulatory associations are ranked by t-score for each cell type. Note that in general, the t-score is only used to compute the p-value and to conclude on the acceptance or rejection of the null hypothesis. Here, p-values are computed to check that the null hypothesis is rejected, yet the t-score is enough for the ranking as the p-value function is monotonic for the t-score when the degree of freedom is constant. The latter is the case, because all the distributions have the same dimensions for one cell type.

We selected the 20 best-ranked regulatory associations of each cell type and called these the *marker regulatory associations*. These associations involve one TF and one gene of interest. We grouped the interactions by gene of interest in the bar plots presented in Fig. 5 and Fig. 6 and we colored the bars by the cell type for which the regulatory association is a marker. The marker regulatory associations are revealing the putative rewiring events in the network, as the same gene of interest can be involved in different regulatory associations for different cell types.

## 3.3 Software implementation

The implementation is in the form of a Python library, compatible with Scanpy. It takes as input an AnnData object containing the scRNA-seq expression matrix, a list of marker genes to analyze, and a list of transcription factors. A visualization tool is provided and allows to reproduce the figures presented in the paper. The experiments can be reproduced by using the repository [https://gitlab.inria.fr/topshap/rewiring\\_detection\\_workshop\\_dmbih\\_2024](https://gitlab.inria.fr/topshap/rewiring_detection_workshop_dmbih_2024)

## 4 Experiments and Results

The experiments have been conducted on two different datasets. The datasets are from human and mouse, and are composed of cells from the immune system.

## 4.1 Evaluation of the results

For each dataset, first a standard Scanpy analysis is performed, including normalizing the expression data, establishing highly-variable genes, computing a PCA, and computing a nearest-neighbor graph. Following this analysis, the population of cells is clustered into cell types using Leiden community detection. Then, by intersecting a species-specific list of known transcription factors to the highly variable genes of the dataset under consideration, we select the highly variable TFs (see Table II). The *rewiring detection method* is applied, using these input data: the raw scRNA-seq dataset, the list of TFs, the list of marker genes that were used to cluster the cells, and the cell types.

Table II: Number of cells and TFs in the datasets.

Dataset	# Transcription factors	# Cells
PBMC	120	2638
Neonatal Mouse	119	2628

The experiments were conducted using the implementation as provided via the URL mentioned above (Section 3.3). The experiments took approximately 4 hours for the PBMC dataset and 2 hours for the Neonatal mouse dataset on a standard laptop with parallelization over eight processors.

## 4.2 PBMC dataset

Fig. 5a projects on a UMAP the different cell types in the dataset identified using marker genes. Fig. 5b show expression levels of each marker gene. Marker genes are known from literature (i.e., other experiments and observations) to be well-expressed in specific cell types, therefore the clusters where they are over-expressed, can be assigned to the corresponding cell type. When marker genes are over-expressed in multiple cell types, several marker genes are needed to fully describe each cell type. In the bar plots of Fig. 5c, marker regulatory associations targeting each marker gene are shown in descending order and the color of the bars represents the cell type for which the regulatory association was detected as statistically different. The heights of each bar correspond to the t-score of the regulatory association.

For example, the association link indicating the regulation of CD8A (Fig. 5c, row 2, 3rd plot) by CEBPB is detected for the CD4 T cells (bar CEBPB, color orange) and the CD8 T cells (bar CEBPB, color green). On the contrary, the regulations of CD8A by LYAR, HOPX, ID2, GTF3A, and JUNB, are unique to CD8 T cells, because the corresponding bars are fully green. This gene also appears to be regulated by MAFB and IRF8 in CD4 T cells (orange). Note that no marker regulatory associations are detected in cell types where CD8A is not expressed (Fig. 5b, row 2, 3rd plot).

The gene LYZ, which is associated to CD14+ monocytes, is also highly expressed in three additional cell types: dendritic cells, megakaryocytes, and FCGR3A+





monocytes. Nonetheless, gene regulation varies depending on the cell type. Only one regulator is detected for two cell types: CEBPB is detected for FCGR3A+ monocytes and CD14+ monocytes. All the others are unique to a cell type. This complex regulation could indicate that the expression of LYZ in different cell types results from its (up)regulation by several TFs.

Interestingly, regulators are not diverse. Among the 160 marker regulatory associations detected, 33 unique TFs are involved. 127 highly-variable TFs were used to build the models and the reduced number of distinct regulators appearing in marker regulatory associations indicates that only a subset of the transcription factors was sufficient to capture important differences of regulation between the cell types. It is worth noting that these are predictions made by a machine learning approach and that the biological importance and validity need to be confirmed. The method “simply” detects differences in the importance of TFs to predict the expression of a TG in a model learned from the data. However, marker regulatory associations targeting a marker gene are often detected for the cell type corresponding to this marker gene. This indicates that the regulatory associations are coherent with respect to current knowledge of cell type marker genes.

For example, the gene MS4A1 is uniquely expressed in the B cells (see Fig. 5b first row, 3rd position). The marker regulatory associations detected involving MS4A1 are almost uniquely found markers for the B cells (see Fig. 5c, first row, 3rd position).

### 4.3 Neonatal mouse dataset

This dataset was initially published by Bacon et al. [35].

Fig. 6a is a UMAP presenting the different cell types identified using the marker genes presented in Fig. 5b. Several individuals composed the original dataset. We selected the data from only one wildtype individual (N705) that represented the best the full dataset.

As previously, the marker regulatory associations targeting each marker gene are ordered in the bar plots of Fig. 5c, and the color of the bars represents the cell type for which the regulatory association was detected as statistically different. The heights of each bar correspond to the t-score of the regulatory association.

In this dataset, four marker genes are implicated in regulatory associations that are in majority detected in one cell type: Cd8b1, Il2ra, Hbb-bs and Itm2a. This detection is reassuring, as the marker genes are mostly expressed in the corresponding cell types (Fig. 6b).

Itm2a displays an interesting pattern, where 14 regulators are detected in Mature T-cells, for which the gene is highly expressed. Nevertheless, a second series of 3 regulators with distinctly lower t-scores are detected for Middle T-cells, where indeed Itm2a is less expressed.

For this dataset, 44 unique TFs were involved in the 100 marker regulatory associations detected, which suggests a larger diversity of TFs than for the PBMC dataset (33 TFs in 160 associations, see above). This could be explained by the fact that here we analysed a cell differen-

tiation trajectory from early to mature T-cells. Developmental processes, such as cell differentiation, are known to involve many TFs. Instead, the PBMC dataset captures mature cell types, where only cell type maintenance is required. It may be that maintenance of a cell type requires fewer TFs than the acquisition of a cell type.

## 5 Conclusion

In this article, we proposed a method to detect the putative rewiring events in a GRN using only a single-cell RNA sequencing dataset. The method relies on computation of machine learning models and local selection of the most important features, using the SHAP framework and computation of confidence intervals. The method was applied to datasets from human and mouse, both composed of cells from the immune system. It showed promising results in terms of detecting rewiring events. Indeed, the marker regulatory associations obtained by our method are coherent with respect to known cell type marker genes. A better assessment of the importance of our method’s results will still require experimental validation. And to ensure reproducible results (at least from the computational point of view), our experiments are available as Jupyter notebooks accompanying this work.

As this method has been designed to be widely applicable, it should be noted that for studies that have other, additional kinds of data available (e.g., single-cell chromatin accessibility data), methods that take into account such data should be favored. Yet, even if we focus on detecting rewiring events with only scRNA-seq data, this field of study is still in its infancy and urgently needs ground truth datasets for better and more detailed benchmarks of old and new methods.

## 6 Acknowledgment

This work was partially funded by PEPR Santé Numérique, project 22-PESN-0002; Fondation ARC under grant ARCPJA22020060002212; Institut National du Cancer under grant PLBIO22-071; the National Institutes of Health (NIH) award R01DC020478 and BQR INSA Lyon 2023 Neurinfo. This work was also supported by the LabEx IMU (ANR-10-LABX-0088). The authors thank J. Luiselli and T. Peyric for their comments.

## References

- [1] S. A. Assi, M. R. Imperato, D. J. Coleman, A. Pickin, S. Potluri, A. Ptasinska, P. S. Chin, H. Blair, P. Cauchy, and S. R. James, “Subtype-specific regulatory network rewiring in acute myeloid leukemia,” *Nature genetics*, vol. 51, no. 1, pp. 151–162, 2019, publisher: Nature Publishing Group US New York.
- [2] N. Bhardwaj, P. M. Kim, and M. B. Gerstein, “Rewiring of Transcriptional Regulatory Networks:



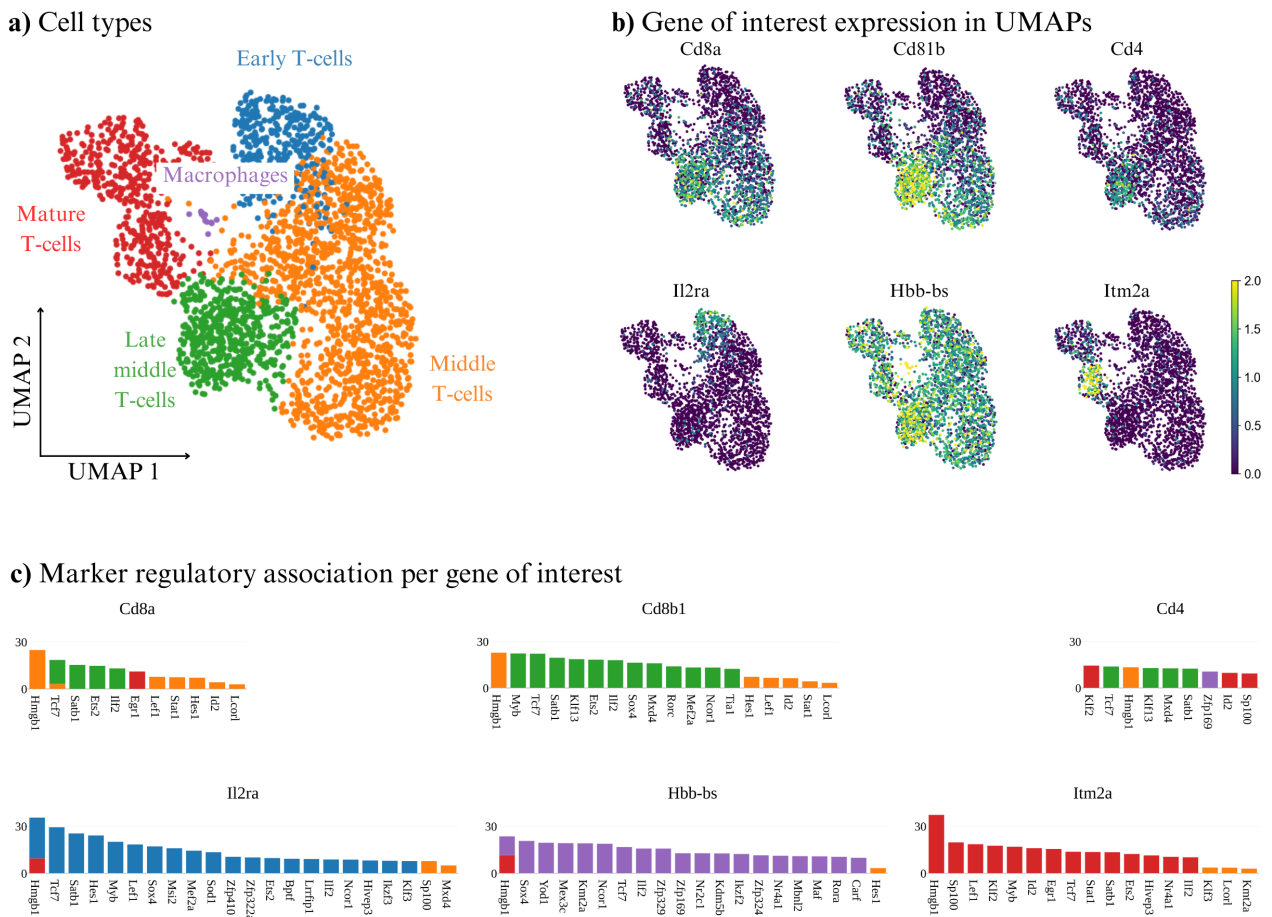


Figure 6: Neonatal mouse dataset. a) The cell types assigned to each group of cells using the marker genes. b) The levels of expression of the marker gene projected on the UMAP. c) The marker regulatory association involving the marker genes. The bar height represents the t-score of the regulatory associations in the cell type corresponding to the color of the bar.

- Hierarchy, Rather Than Connectivity, Better Reflects the Importance of Regulators,” *Science Signaling*, vol. 3, no. 146, pp. ra79–ra79, Nov. 2010, publisher: American Association for the Advancement of Science.
- [3] D. Srivastava and S. Mahony, “Sequence and chromatin determinants of transcription factor binding and the establishment of cell type-specific binding patterns,” *Biochimica Et Biophysica Acta. Gene Regulatory Mechanisms*, vol. 1863, no. 6, p. 194443, Jun. 2020.
- [4] W. A. Whyte, D. A. Orlando, D. Hnisz, B. J. Abraham, C. Y. Lin, M. H. Kagey, P. B. Rahl, T. I. Lee, and R. A. Young, “Master transcription factors and mediator establish super-enhancers at key cell identity genes,” *Cell*, vol. 153, no. 2, pp. 307–319, 2013, publisher: Elsevier.
- [5] R. Baumstark, S. Hänzelmann, S. Tsuru, Y. Schaerli, M. Francesconi, F. M. Mancuso, R. Castelo, and M. Isalan, “The propagation of perturbations in rewired bacterial gene networks,” *Nature Communications*, vol. 6, no. 1, p. 10105, Dec. 2015, publisher: Nature Publishing Group.
- [6] S. Sarda and S. Hannenhalli, “High-Throughput Identification of Cis-Regulatory Rewiring Events in Yeast,” *Molecular Biology and Evolution*, vol. 32, no. 12, pp. 3047–3063, Dec. 2015.
- [7] T. L. Davis and I. Rebay, “Master regulators in development: Views from the Drosophila retinal determination and mammalian pluripotency gene networks,” *Developmental Biology*, vol. 421, no. 2, pp. 93–107, Jan. 2017.
- [8] A. N. Brooks, S. Turkarslan, K. D. Beer, F. Yin Lo, and N. S. Baliga, “Adaptation of cells to new environments,” *WIREs Systems Biology and Medicine*, vol. 3, no. 5, pp. 544–561, Sep. 2011.
- [9] A. Califano, “Rewiring makes the difference,” *Molecular Systems Biology*, vol. 7, no. 1, p. 463, Jan. 2011, publisher: John Wiley & Sons, Ltd.
- [10] G. Karlebach and R. Shamir, “Modelling and analysis of gene regulatory networks,” *Nature Reviews Molecular Cell Biology*, vol. 9, no. 10, pp. 770–780, Oct. 2008, publisher: Nature Publishing Group.
- [11] H. Han, H. Shim, D. Shin, J. E. Shim, Y. Ko, J. Shin, H. Kim, A. Cho, E. Kim, and T. Lee, “TRRUST: a reference database of human transcriptional regulatory interactions,” *Scientific reports*, vol. 5, no. 1, p. 11432, 2015, publisher: Nature Publishing Group UK London.
- [12] N. Bonhoure, G. Bounova, D. Bernasconi, V. Praz, F. Lammers, D. Canella, I. M. Willis, W. Herr, N. Hernandez, M. Delorenzi, T. C. Consortium, N. Hernandez, M. Delorenzi, B. Deplancke, B. Desvergne, N. Guex, W. Herr, F. Naef, J. Rouge-mont, U. Schibler, T. Andersin, P. Cousin, F. Gilar-di, P. Gos, F. Lammers, S. Raghav, D. Villeneuve, R. Fabbretti, V. Vlegel, I. Xenarios, E. Migliavacca, V. Praz, F. David, Y. Jarosz, D. Kuznetsov, R. Liechti, O. Martin, J. Delafontaine, J. Cajan, K. Gustafson, I. Krier, M. Leleu, N. Molina, A. Naldi, L. Rib, L. Symul, and G. Bounova, “Quantifying ChIP-seq data: a spiking method providing an internal reference for sample-to-sample normalization,” *Genome Research*, vol. 24, no. 7, pp. 1157–1168, Jul. 2014, company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.
- [13] T. Xu, L. Ou-Yang, X. Hu, and X.-F. Zhang, “Identifying Gene Network Rewiring by Integrating Gene Expression and Gene Network Data,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 15, no. 6, pp. 2079–2085, Nov. 2018, conference Name: IEEE/ACM Transactions on Computational Biology and Bioinformatics.
- [14] D. Tian, Q. Gu, and J. Ma, “Identifying gene regulatory network rewiring using latent differential graphical models,” *Nucleic Acids Research*, vol. 44, no. 17, p. e140, Sep. 2016.
- [15] P. Badia-i Mompel, L. Wessels, S. Müller-Dott, R. Trimbour, R. O. Ramirez Flores, R. Argelaguet, and J. Saez-Rodriguez, “Gene regulatory network inference in the era of single-cell multi-omics,” *Nature Reviews Genetics*, vol. 24, no. 11, pp. 739–754, 2023, publisher: Nature Publishing Group UK London.
- [16] T. Moerman, S. Aibar Santos, C. Bravo González-Blas, J. Simm, Y. Moreau, J. Aerts, and S. Aerts, “GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks,” *Bioinformatics (Oxford, England)*, vol. 35, no. 12, pp. 2159–2161, Jun. 2019.
- [17] S. M. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” in *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc., 2017.
- [18] M. T. Ribeiro, S. Singh, and C. Guestrin, “‘Why Should I Trust You?’: Explaining the Predictions of Any Classifier,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco California USA: ACM, Aug. 2016, pp. 1135–1144.
- [19] C. B. González-Blas, S. De Winter, G. Hulselmans, N. Hecker, I. Matetovici, V. Christiaens, S. Poo-vathingal, J. Wouters, S. Aibar, and S. Aerts, “SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks,” Aug. 2022.

- [20] A. G. Anderson, B. B. Rogers, J. M. Loupe, I. Rodriguez-Nunez, S. C. Roberts, L. M. White, J. N. Brazell, W. E. Bunney, B. G. Bunney, S. J. Watson, J. N. Cochran, R. M. Myers, and L. F. Rizzardi, “Single nucleus multiomics identifies ZEB1 and MAFB as candidate regulators of alzheimer’s disease-specific cis-regulatory elements,” *Cell Genomics*, vol. 3, no. 3, p. 100263, 2023-03.
- [21] C. Burdziak, E. Azizi, S. Prabhakaran, and D. Pe’er, “A nonparametric multi-view model for estimating cell type-specific gene regulatory networks,” *arXiv e-prints*, pp. arXiv-1902, 2019.
- [22] P. Bachiredy, E. Azizi, C. Burdziak, V. N. Nguyen, C. S. Ennis, K. Maurer, C. Y. Park, Z.-N. Choo, S. Li, S. H. Gohil, N. G. Ruthen, Z. Ge, D. B. Keskin, N. Cieri, K. J. Livak, H. T. Kim, D. S. Neuberg, R. J. Soiffer, J. Ritz, E. P. Alyea, D. Pe’er, and C. J. Wu, “Mapping the evolution of t cell states during response and resistance to adoptive cellular therapy,” *Cell Reports*, vol. 37, no. 6, p. 109992, 2021-11.
- [23] P. Wang, X. Wen, H. Li, P. Lang, S. Li, Y. Lei, H. Shu, L. Gao, D. Zhao, and J. Zeng, “Deciphering driver regulators of cell fate decisions from single-cell transcriptomics data with CEFCON,” *Nature Communications*, vol. 14, no. 1, p. 8459, 2023-12-20.
- [24] D. Thompson, A. Regev, and S. Roy, “Comparative analysis of gene regulatory networks: From network reconstruction to evolution,” *Annual Review of Cell and Developmental Biology*, vol. 31, no. 1, pp. 399–428, 2015-11-13.
- [25] J. K. Pritchard, M. Stephens, and P. Donnelly, “Inference of population structure using multilocus genotype data,” *Genetics*, vol. 155, no. 2, pp. 945–959, 2000, publisher: Oxford University Press.
- [26] S. Zhang, S. Pyne, S. Pietrzak, S. Halberg, S. G. McCalla, A. F. Siahpirani, R. Sridharan, and S. Roy, “Inference of cell type-specific gene regulatory networks on cell lineages from single cell omic datasets,” *Nature Communications*, vol. 14, no. 1, p. 3064, 2023-05-27.
- [27] T. Xu, L. Ou-Yang, H. Yan, and X.-F. Zhang, “Time-varying differential network analysis for revealing network rewiring over cancer progression,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 4, pp. 1632–1642, 2021-07-01.
- [28] Z. Duren, W. S. Lu, J. G. Arthur, P. Shah, J. Xin, F. Meschi, M. L. Li, C. M. Nemeč, Y. Yin, and W. H. Wong, “Sc-compReg enables the comparison of gene regulatory networks between conditions using single-cell data,” *Nature Communications*, vol. 12, no. 1, p. 4763, 2021-08-06.
- [29] J. Xie, Y. Yin, F. Yang, J. Sun, and J. Wang, “Differential network analysis reveals regulatory patterns in neural stem cell fate decision,” *Interdisciplinary Sciences: Computational Life Sciences*, vol. 13, no. 1, pp. 91–102, 2021-03.
- [30] S. Lou, T. Li, X. Kong, J. Zhang, J. Liu, D. Lee, and M. Gerstein, “TopicNet: a framework for measuring transcriptional regulatory network change,” *Bioinformatics*, vol. 36, pp. i474–i481, 2020-07-01.
- [31] F. A. Wolf, P. Angerer, and F. J. Theis, “SCANPY: large-scale single-cell gene expression data analysis,” *Genome Biology*, vol. 19, no. 1, p. 15, Dec. 2018.
- [32] U. Faigle and W. Kern, “The shapley value for cooperative games under precedence constraints,” *International Journal of Game Theory*, vol. 21, pp. 249–266, 1992.
- [33] E. Štrumbelj and I. Kononenko, “Explaining prediction models and individual predictions with feature contributions,” *Knowledge and Information Systems*, vol. 41, no. 3, pp. 647–665, 2014-12.
- [34] O. J. Dunn, “Multiple Comparisons among Means,” *Journal of the American Statistical Association*, vol. 56, no. 293, pp. 52–64, Mar. 1961, publisher: Taylor & Francis.
- [35] W. A. Bacon, R. S. Hamilton, Z. Yu, J. Kieckbusch, D. Hawkes, A. M. Krzak, C. Abell, F. Colucci, and D. S. Charnock-Jones, “Single-Cell Analysis Identifies Thymic Maturation Delay in Growth-Restricted Neonatal Mice,” *Frontiers in Immunology*, vol. 9, p. 2523, 2018.