様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　６ 年　 ５ 月 ２８ 日現在

機関番号：１２６０１
研究種目：研究活動スタート支援
研究期間：2022～2023
課題番号：２２Ｋ２１３０１

研究課題名（和文）Inferring cells differentiation processes from single-cell Multiome ATACseq+ RNAseq data.

研究課題名（英文）Inferring cells differentiation processes from single-cell Multiome ATACseq+ RNAseq data.

研究代表者

Ｌｏｚａ　Ｌｏｐｅｚ．Ｍａｒｔｉｎ．ｄｅ．Ｊｅｓｕｓ（LOZA LOPEZ, MARTIN DE JESUS）

東京大学・医科学研究所・助教

研究者番号：３０９６３１９４

交付決定額（研究期間全体）：（直接経費）　　2,200,000円

研究成果の概要（和文）：私たちは、マクロファージと呼ばれる免疫細胞の遺伝子挙動を理解するためのモデルを開発しました。変化の履歴を保持する遺伝子（ヒステリシス遺伝子）に焦点を当てることで、細胞の再プログラミング中にこれらの遺伝子がどのように振る舞うかを予測する方法を作成しました。私たちのモデルは遺伝子活動を正確にシミュレートし、免疫細胞がどのように適応するかを探るのに役立ちます。この研究は、マクロファージに関連する免疫応答の理解を深め、より良いがん治療法の開発につながる可能性があります。

研究成果の学術的意義や社会的意義
Our research helps us understand how immune cells called macrophages change and adapt. This knowledge can lead to better cancer treatments by improving how we predict and influence immune responses. Ultimately, it can enhance our ability to fight diseases and improve public health.

研究成果の概要（英文）：We developed a model to understand gene behavior in immune cells called macrophages. By focusing on genes that preserved a history of their changes (hysteresis genes), we created a method to predict how these genes behave during cell reprogramming. Our model accurately simulated gene activity, helping us explore how immune cells adapt. This research could improve our understanding of immune responses related to macrophages and lead to better cancer treatments.

研究分野：Bioinformatics

キーワード：RNA modeling　Cells reprogramming　Simulation　Macrophages　Immunology

１．研究開始当初の背景 (Background at the start of research)

The advances in single-cell sequencing techniques have opened up new horizons for complex cell analysis, with a profound impact. A significant breakthrough in this field was the identification of inherent changes in gene expression, such as RNA velocity and cell differentiation lineage. However, these methods often fall short in capturing the dynamics of cells in complex scenarios due to their simplistic assumptions. In this research, I propose a novel approach that incorporates single-cell multiomics data to address this limitation. By harnessing the power of fuzzy logic, gene regulatory networks, and information from scRNA + scATAC data, I aim to develop a computational method that can analyze the joint effects of transcription factors and gene expression in defining cell types and lineage processes. This innovative methodology could revolutionize exploratory analysis and in-silico perturbation and prediction of cell states, thereby enhancing cell reprogramming in gene therapy.

２．研究の目的 (Research objectives)

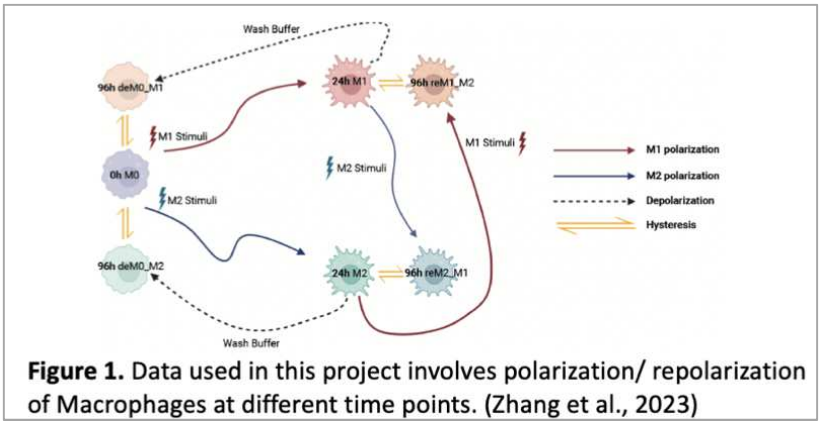

Figure 1. Data used in this project involves polarization/ repolarization of Macrophages at different time points. (Zhang et al., 2023)

During the course of this research, several similar methods, such as CellOracle, were published in renowned journals. Then, to improve the impact of our work, we focused on a specific problem: the cell reprogramming scenario in macrophages. In a paper we published during this research (Zhang Y. et al., 2023), we investigated a set of genes displaying a "hysteresis" profile during the reprogramming of macrophages from M0 to M1 or M2 and vice versa (Liu S.X. et al., 2020). Our focus shifted to designing a fuzzy logic model to estimate expression levels in macrophages, specifically targeting a set of hysteresis genes, such as those that retain an M1 profile after repolarization to M0.
Figure 1 depicts the data used in this research, which involves polarization and repolarization data of M0, M1, and M2 macrophages at different time points.
The revised objectives of this project are:

(1) To develop a fuzzy logic model for simulating the gene expression of hysteresis genes in macrophages.

(2) To simulate expression levels of hysteresis genes at time points for which data is not available.

３．研究の方法 (Research method)

Figure 2 illustrates the workflow for estimating the expression levels of hysteresis genes during macrophage polarization. This workflow is similar to our original proposal but excludes gene regulatory networks. The steps are as follows:

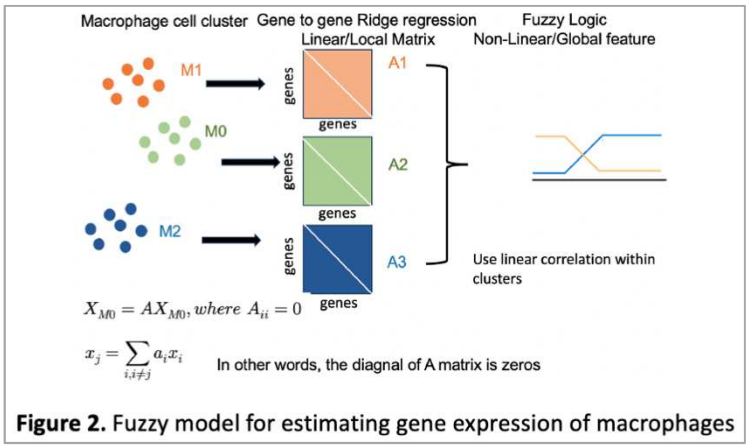

Figure 2. Fuzzy model for estimating gene expression of macrophages

(1) Definition of clusters: Clusters are defined in the data, with the granularity adjusted according to the heterogeneity of the samples. In this case, as we are analyzing sorted macrophage cells, we set a low granularity.

(2) Training linear models: For each cluster, we train a linear model to estimate gene expression based on the expression levels of other genes. This can be viewed as a gene-gene model, where the expression level of a specific gene depends on the expression levels of other genes. In this model, the diagonal is set to zero to avoid self-dependence during training.

(3) Merging models using fuzzy logic: Using the trained linear models for each cluster, we create a final model by merging them using fuzzy logic rules. The distance between clusters is used as a variable to guide the interpolation between models. Ultimately, each cell will have its own gene-gene matrix according to its distance from the cluster centers.

(4) Simulating gene expression changes: The obtained model is then used to simulate changes in the expression of hysteresis genes.

4. 研究成果 (Research results)

(1) Training linear models: I successfully trained the linear models for each group with small errors. A key parameter to consider during training is lambda, a parameter in Ridge regression that "controls" the number of important features (genes in our case) related to the simulated expression levels. Setting a low lambda parameter results in lower estimation error but could produce a complex network of gene-gene relationships, complicating the final analysis. Conversely, a high lambda reduces the number of gene-gene interactions, making analysis easier but with higher prediction errors. After comparing different lambda values, we chose a high lambda to obtain a reduced but significant number of gene-gene interactions for each hysteresis

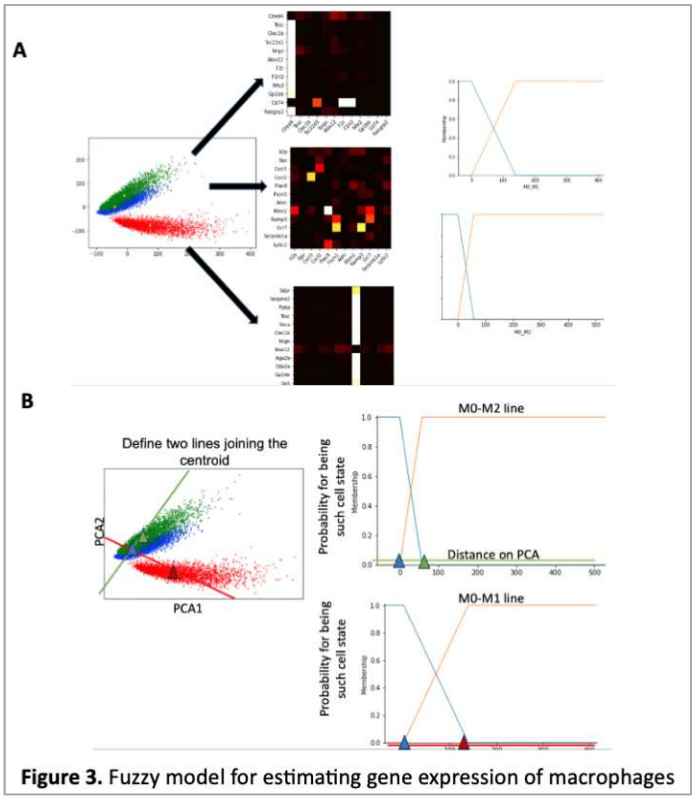

Figure 3. Fuzzy model for estimating gene expression of macrophages

gene.

(2) Creating cluster-specific models: After setting the lambda parameter, we obtained a linear model for each macrophage cluster. Figure 3A shows the PCA representation of the macrophage clusters and an example of the gene-gene matrix obtained for each; only the most significant interactions are shown for simplicity.

(3) Assembling the final model: I assembled the final model using the cluster-specific gene-gene interaction matrices, guided by the distance of cluster centroids in the PCA space for fuzzy interpolation. Figure 3 B shows an example of this interpolation.

(4) Testing the model: Once the fuzzy logic model was assembled, we tested it by predicting gene expression levels 6 hours after repolarization to M1. It is important to note that this dataset was not used for training, eliminating any bias towards the training dataset. Figure 4 shows the comparison of simulated versus real gene expression levels for a set of hysteresis genes. Some genes showed a high similarity between the simulated and real expression distributions (Ccl3 and Cxcl9), while others showed low agreement, indicating potential improvements needed in the modeling (Ccl5 and Cxcl10).

In summary, we trained a fuzzy linear model to predict the gene expression levels of hysteresis genes in macrophages after repolarization from M1 to M0. The model successfully reproduced the observed distribution of transcription levels when compared with an unseen dataset, highlighting its potential for investigating dynamic changes in cells. While some gene distributions differed from expectations, as shown in Figure 4B, we view this as an opportunity for future improvements to achieve more accurate cell dynamics modeling.

These results will be included in a subsequent publication of our previous findings, which we expect to be ready for publication later this year.

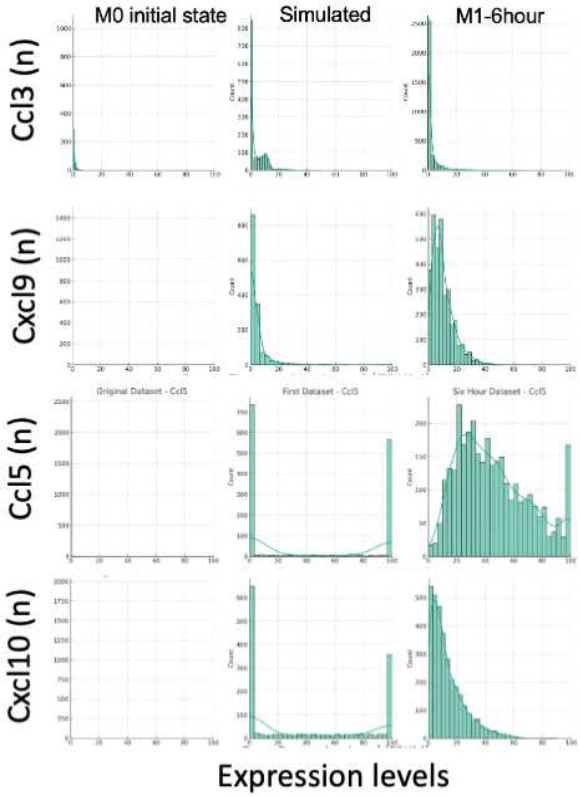

Figure 4. Simulated vs. observed expression levels of hysteresis genes in the M1 6hrs dataset.

５．主な発表論文等

〔雑誌論文〕　計3件（うち査読付論文　3件／うち国際共著　3件／うちオープンアクセス　3件）

１．著者名 Loza Martin、Vandenbon Alexis、Nakai Kenta	４．巻 52
２．論文標題 Epigenetic characterization of housekeeping core promoters and their importance in tumor suppression	５．発行年 2023年
３．雑誌名 Nucleic Acids Research	６．最初と最後の頁 1107～1119
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/nar/gkad1164	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Garcia M.、Melgoza E. A.、Hernandez S.、Hinojosa D.、Resendiz M.、Llamas M. A.、Loza M.、Valenzuela O.、Soto A.、Hernandez M. A.、Mata V.、Cassaniti I.、Sammartino J. C.、Ferrari A.、Simonelli L.、Pedotti M.、Sun R.、Zuo F.、Baldanti F.、Varani L.、Marcotte H.、Pan-Hammarstrom Q.、Hernandez J.	４．巻 26
２．論文標題 19n01, a broadly neutralizing antibody against omicron BA.1, BA.2, BA.4/5, and other SARS-CoV-2 variants of concern	５．発行年 2023年
３．雑誌名 iScience	６．最初と最後の頁 106562～106562
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.isci.2023.106562	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Jiravejchakul Natnicha、Abe Gabriela L.、Loza Martin、Park Soyoung、Matangkasombut Ponpan、Sasaki Jun-Ichi、Imazato Satoshi、Diez Diego、Standley Daron M.	４．巻 24
２．論文標題 Intercellular crosstalk in adult dental pulp is mediated by heparin-binding growth factors Pleiotrophin and Midkine	５．発行年 2023年
３．雑誌名 BMC Genomics	６．最初と最後の頁 1-10
掲載論文のDOI（デジタルオブジェクト識別子） 10.1186/s12864-023-09265-w	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

〔学会発表〕　計9件（うち招待講演　1件／うち国際学会　7件）

１．発表者名 MARTIN LOZA, ALEXIS VANDENBON, and KENTA NAKAI
２．発表標題 Human housekeeping cis-regulatory elements and their involvement in tumor suppression
３．学会等名 Indian Conference on Bioinformatics（招待講演）
４．発表年 2023年

１．発表者名
MARTIN LOZA, ALEXIS VANDENBON, and KENTA NAKAI

２．発表標題
Epigenetic characterization of human housekeeping cis-regulatory elements and their implications in cancer development

３．学会等名
International Workshop on Advanced Genomics (14AWG) (国際学会)

４．発表年
2023年

１．発表者名
MARTIN LOZA, ALEXIS VANDENBON, and KENTA NAKAI

２．発表標題
Characterization of housekeeping regulatory elements in the human genome

３．学会等名
CSHL Meeting: The Biology of Genomes (国際学会)

４．発表年
2023年

１．発表者名
MARTIN LOZA, ALEXIS VANDENBON, and KENTA NAKAI

２．発表標題
Unveiling the existence of human housekeeping cis-regulatory elements: epigenetic characterization and implications in cancer development

３．学会等名
INCOB 2023

４．発表年
2023年

１．発表者名
Loza Martin

２．発表標題
Accurate integration of single cell transcriptome replicates

３．学会等名
2022 HCA Latin America Symposium (国際学会)

４．発表年
2022年

１．発表者名
Loza Martin

２．発表標題
Housekeeping enhancers in the human genome

３．学会等名
The International Conference on Bioinformatics (InCoB)（国際学会）

４．発表年
2022年

１．発表者名
Loza Martin

２．発表標題
Unbiased integration of single cell transcriptome replicates

３．学会等名
Simposio de Estudiantes Hispanoablantes de bioinformatica y Biologia Computacional (SEHbioinfo))（国際学会）

４．発表年
2022年

１．発表者名
Loza Martin

２．発表標題
Characterization of housekeeping enhancers and their targets in the human genome

３．学会等名
NGS EXPO IFReC（国際学会）

４．発表年
2022年

１．発表者名
Loza Martin

２．発表標題
Unbiased integration of single cell transcriptome replicates

３．学会等名
IIBMP（国際学会）

４．発表年
2022年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関