
Linear programming-based prediction of immune cell composition

SeongRyeol Moon

Department of Health Sciences and Technology (GAIHST),
Gachon University,
Incheon 21999, South Korea
Email: moonsr1982@gmail.com

Junbae Oh

Department of Software Engineering,
Ajou University,
Suwon 16499, Gyeonggi-do, South Korea
Email: dhwnsqo1@ajou.ac.kr

Seungyeon Nam*

Department of Life Sciences,
Gachon University, Seongnam 13120, South Korea
and
College of Medicine,
Gachon University,
Incheon 21565, South Korea
Email: nams@gachon.ac.kr
*Corresponding author

Abstract: Immune cell types play critical roles in pathogenesis of loss of tolerance, resulting in autoimmune disorders. Immune cell composition has now been recognised as a clinical tool for monitoring disease progression, as typically inferred from gene expression information of a large prior set of differentiation genes, demarcating specific immune cell types. However, given a small set of gene families, immune cell type inference has not been established. Here we used Linear Programming (LP), by a Simplex method, to infer fractions of immune cells, based on a small set of genes used in cell surface marker experiments. To evaluate the accuracy of our method, we created multiple simulated data sets, and evaluated their performance against Digital Cell Quantisation (DCQ) and ImmQuant. Finally, we applied LP method to real biological data, from multiple Systemic Lupus Erythematosus (SLE) patients, versus healthy controls, to inspect compositional changes of immune cell types, in SLE pathogenesis.

Keywords: linear programming; immune cells; immune cell type prediction

Reference to this paper should be made as follows: Moon, S., Oh, J. and Nam, S. (2020) 'Linear programming-based prediction of immune cell composition', *Int. J. Data Mining and Bioinformatics*, Vol. 23, No. 2, pp.176–187.

Biographical notes: SeongRyeol Moon received his MS degree in Science from Gachon University, South Korea, in 2019. Currently, he is working toward the PhD degree in the Department of Health Sciences and Technology, Gachon Advanced Institute for Health Sciences and Technology (GAIHST), Gachon University, Incheon, South Korea. His current research interests include network analysis, machine-learning application in biological big data, and next-generation sequencing (NGS) analysis in clinical samples.

Junbae Oh received his BSc degree in Computer Science from Ajou University, South Korea. His current research interests include systems biology.

Seungyeon Nam received his PhD degree in Bioinformatics from Seoul National University in 2008. From 2011 to 2015, he worked as a Senior Researcher in New Experimental Therapeutics (NExT) Branch in National Cancer Centre, South Korea. Currently, he is an Assistant Professor at the College of Medicine of Gachon University, South Korea. His research area includes systems biology, non-coding RNAs, and druggable genome for various diseases.

This paper is a revised and expanded version of the workshop paper entitled "Predicting Immune Cell Composition Using Linear Programming" presented at IEEE International Conference on Bioinformatics and Biomedicine (IEEE BIBM), San Diego, USA, 18–21. Nov. 2019.

1 Introduction

In the immune system, diverse cell types play essential roles during differentiation, disease development, and various metabolism-specific environments (Hobson-Gutierrez and Carmona-Fontaine, 2018; Kim et al., 2020). In fact, under diverse conditions, immune cells phenotypically change (Hobson-Gutierrez and Carmona-Fontaine, 2018), and differentiate into to other cell type lineages (Brodin and Davis, 2017; Kim et al., 2020). In humans, the immune system constitution indicates immune cell subpopulations, and their distinctive protein profiles, which affect individual health statuses (Brodin and Davis, 2017; Kaczorowski et al., 2017). Such distinct phenotypes, and cell types in the immune system, have been delineated by flow cytometry (Chattopadhyay and Roederer, 2015), mass cytometry (Bandura et al., 2009), and single-cell RNA-sequencing (scRNA-seq) (Macosko et al., 2015).

Recently, the composition of immune cells has been recognised as an emerging tool (Macosko et al., 2015) for predicting clinical outcomes of major diseases, including tumours (Lyons et al., 2017). To determine *in vivo* immune cell compositions, tissue samples are, in general, required. To date, typical next-generation sequencing for transcriptomics, usually referred to as bulk RNA-seq, has been performed on tissue samples (Kim et al., 2020). However, because tissues are heterogeneous for different developmental states involving a diverse variety of cell types, the observed transcriptomic profile may reflect an amalgamation of subpopulations (Kim et al., 2020), precluding delineation of individual cell types.

Presently, scRNA-seq has been shown to be a promising and emerging approach for immune cell type studies (Han et al., 2014; Macosko et al., 2015). However there are

many bulk RNA-seq immune cell studies already deposited in public repositories, and due to the cost of scRNA-Seq, bulk RNA-seq remains in widespread use. Thus, with regard to deposited bulk RNA-Seq data sets, methodologies for deconvolution are still demanded. Since immune cell types can be discriminated due to their differential gene expression signatures, including cell surface marker gene products (Lyons et al., 2017), deconvolution tools can now be used to infer immune cell fractions (Altboum et al., 2014; Frishberg et al., 2016; Sturm et al., 2019). Deconvolution is useful for a large set of genes, and gene expression profiling data for each immune cell. To that end, less than a dozen markers are used in cell surface marker experiments (typically, Fluorescence-Activated Cell Sorting, FACS) to experimentally quantify immune cells. So, given a small set of genes, a working model for the inference of immune cell fractions is required. In that regard, the objective of the algorithm introduced below is to predict cell type composition by Linear Programming (LP), based on the expression of marker genes. Finally, we applied our LP method to assess data from a widespread autoimmune disease, Systemic Lupus Erythematosus (SLE) (Park and Nam, 2019), to determine whether or not our algorithm can find immune cell composition changes.

2 Methods

Immune cell composition inference by LP: Given expression of cell surface marker genes in immune cells, we aimed to infer immune cell composition from gene expression profiles, for a specific sample.

To that end, gene expression data (“ E ”) of immune cell mixture samples was represented in a matrix, in which rows and columns indicate maker genes ($m = 1, \dots, k$) and samples ($s = 1, \dots, u$) (Figure 1(a), input, left). We then incorporated prior knowledge for cell surface marker immunophenotypes, in immune cell types Figure 1(b) to predict immune cell composition by our algorithm. Similarly, prior knowledge (“ R ”) was represented as a binary matrix, in which rows and columns indicate marker genes ($m = 1, \dots, k$) and immune cell types ($i = 1, \dots, c$) (Figure 1(b), right). From matrix E (expression), all the pairwise correlations between two genes were represented as a Kernel matrix (matrix K) (Figure 1(a), output, right). Moreover, each element in the matrix K was obtained by Radial Basis Function (RBF). Matrix K , therefore, represents a relationship between marker genes of immune cell types. In other word, element $K(e_i, e_j)$, in matrix K , indicates RBF between the row vectors of two immune cell types i and j in matrix E .

$$K(e_i, e_j) = \exp\left(-\frac{e_i - e_j^2}{2\sigma^2}\right) \quad (1)$$

where

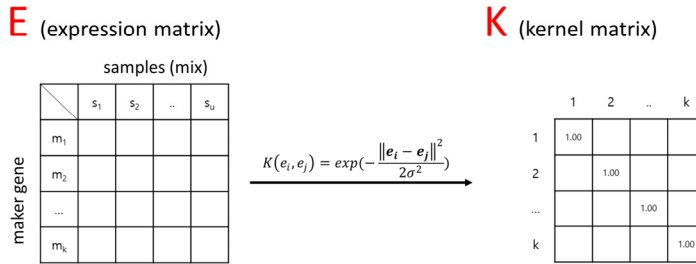
e_k = the row vector of marker gene k in matrix E

σ = free parameter (generally 1)

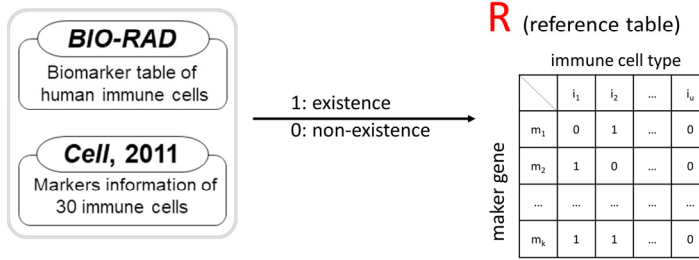
The column vector, x , for an immune cell mixture sample, indicates the proportions of immune cell types, and each element x_i ($i = 1, \dots, n$), in x , indicates the i -th immune cell type. The column vector, y , indicates gene expression for a given sample (equivalently, column) in matrix E .

Figure 1 Preparation for immune cell composition inference. (a) Kernel matrix (K) represents all pairwise correlations between each gene of matrix E . (b) The reference table was created by integrated the markers of immune cells identified in previous studies with those provided by BIO-RAD. The reference table R is represented by a binary matrix. (c) The coefficients matrix (A) is generated using the reference table (R) and the kernel matrix (K)

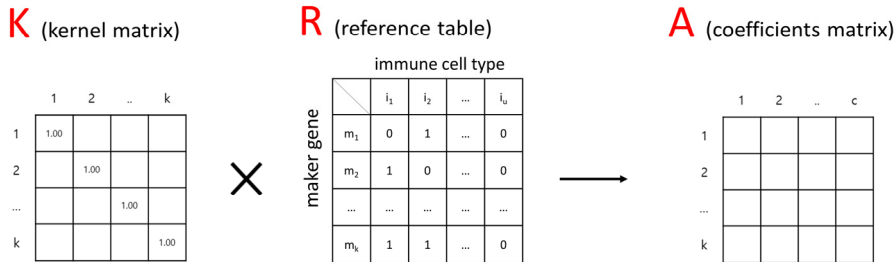
a.



b.



c.



Our purpose here was to infer immune cell proportions, in each sample, in matrix E . To solve this problem, we set up linear programming, by a Simplex method, as follows:

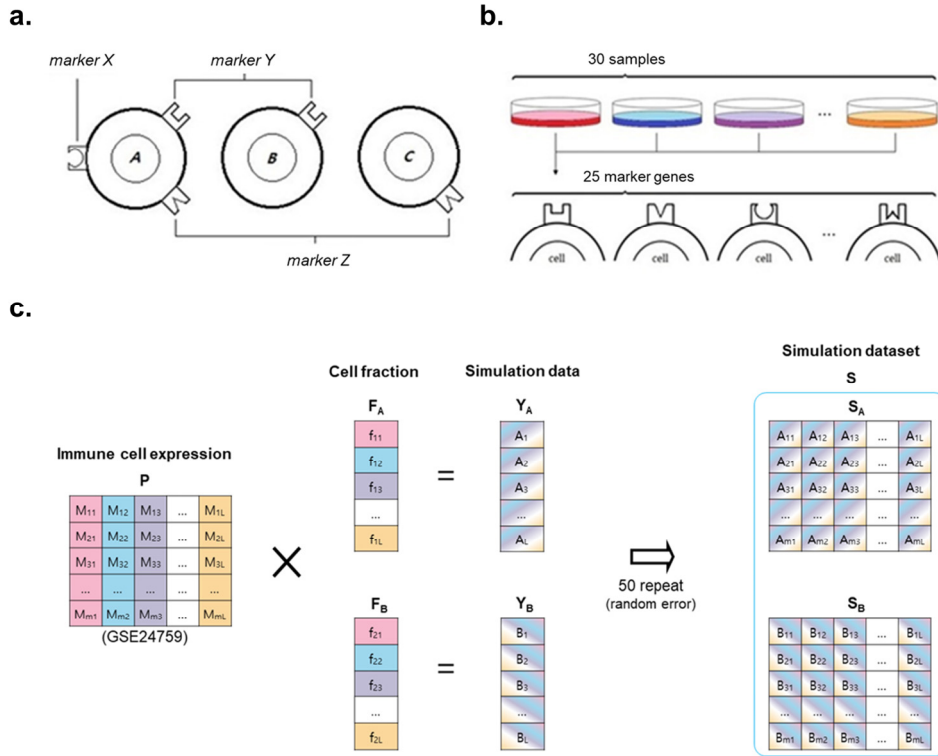
$$\text{Maximise } c^T x \quad (2)$$

Subject to $Ax=y$ and $x \geq 0$, where $A = KR$.

In the problem, the column vector, c , was set to all entries ones, and ^{T} is the matrix transpose. Please refer schematic diagram of matrix A to Figure 1(c).

Simulation data generation: To estimate the accuracy of the algorithm for predicting cell type composition, we constructed a simulated data set from a well-established gene expression data set of purified immune cell types (see Figure 2) (Novershtern et al., 2011). Among the 38 immune cells classified, according to gene expression profiling, 30 immune cells were selected, except stem cells and difficult-to-characterise immune cells. Additionally, 25 cell surface markers, used for profiling, were selected as marker genes. The expression levels of such marker genes were then used as mean values (see Figures 2(a) and 2(b)).

Figure 2 Overview of simulations. (a) Schemes of cell surface markers of immune cells. Cell surface markers are unique or shared. (b) 25 marker genes were selected from 30 immune cell types. (c) Simulation data set generation. The simulated data set was constructed from well-established gene expression data of purified immune cell types



In addition, the simulated dataset consisted of two classes, A and B . Say the marker gene expression matrix of L immune cell types ($l=1, \dots, L$), purified or isolated by Fluorescence-Activated Cell Sorting (FACS), as a matrix P . In the simulation, each sample belonging to class κ ($k \in \{A, B\}$) was generated by mixing gene expression values, of L immune cell types ($l=1, \dots, L$), in matrix P . In a sample belonging to class A , all the genes were mixed in equal amounts throughout the L immune cell types. But, in

a class B sample, the only subset of all the genes was negatively or positively affected by an effect size, in comparison with class A , while the other genes were not affected. Each class had 50 generated samples. Therefore, the detailed scheme for each sample is described below. Let's assume synthesised gene expression y_{jk} for gene j ($j = 1, \dots, m$) in a sample belonging to the class k ($k \in \{A, B\}$), is:

$$y_{jk} = \sum_{i=1 \dots L} b_{ji} \cdot f_{ik} + \varepsilon_{jk} \quad (3)$$

In other words, y_{jk} was generated by mixing gene expression values (b_{ji}) of gene j from $i=1$ to L immune cell types, in numbers represented by f_{ik} 's. The b_{ji} 's were obtained from known gene expression information(GEO) (GEO accession: GSE24759) (Novershtern et al., 2011), for the purified L immune cell types (matrix P). Thus, f_{ik} was the fraction of the i -th immune cell type, over all the immune cell types, depending on k . The variable ε_{jk} is an error term, while the variable c_{ik} indicates which immune cell types were affected by the effect size s , and f_{ik} , ε_{jk} , and c_{ik} are expressed as below. Depending on the index i , when $i \in \{U, D\}$, the c_{ik} of an immune cell type is affected (e.g., addition, subtraction) by effect size s (setting to 0.1). When $i \notin \{U, D\}$, the effect size was not considered.

$$f_{ik} = \frac{c_{ik}}{\sum_{i=1 \dots L} c_{ik}} + \varepsilon_{ik} \quad (4)$$

$$\varepsilon_{ik} = N(0, \sigma_i), \sigma_i = \gamma_c \cdot c_{ik} / \sum_{i=1 \dots L} c_{ik} \quad (5)$$

where $\gamma_c = 0.01$

$$c_{ik} = \begin{cases} 1/L, & \text{when } k = A \\ 1/L, & \text{when } k = B, i \notin \{U, D\} \\ (1+s)/L, & \text{when } k = B, i \in \{U, D\} \\ (1-s)/L, & \text{when } k = B, i = D \end{cases} \quad (6)$$

For example, setting L to 10, with gene expression of 10 different immune cell types from GSE24759 selected, with the number of marker genes, for the 10 cell types, m set to 25. So, the synthetic dataset, containing 50 class A samples (henceforth, matrix S_A), and 50 class B samples (henceforth, matrix S_B) were generated (i.e., a 25 by 100 matrix; henceforth, matrix $S = [S_A | S_B]$) (see Figure 2(c)). In that instance, the number of cell types belonging to U and D was two each.

For simplicity in performance evaluation, the samples in matrix S were assumed to be paired, resulting in 50 total pairs. The pairs were between matrixes; for example, the first sample in matrix S_A was paired with the first one in matrix S_B (Figure 2(c), rightmost panel).

Given a pair of classes A and B samples, for each immune cell type, (i) when f_{ik} of an immune cell type in the class B sample belongs to U , the composition difference of the

cell type (i.e., composition of class B sample minus that of class A sample) was labelled as positive (the first and the three lines in equation (6)); (ii) when f_{ik} of an immune cell type, in the sample class B , belonged to D , the composition difference of the cell type (i.e., the composition of class B sample minus that of the class A sample) was labelled as negative (the first and the fourth lines in equation (6)); and (iii) when c_{ik} of an immune cell type, in the sample class B , did not belong to $\{U, D\}$, the composition difference of the cell type (i.e., composition of class B sample minus that of class A sample) was labelled as no difference (the first and the second lines in equation (6)).

In each pair, we calculated L , the number of immune cell types, as composition differences, between classes A and B . Thus, for the 50 pairs, a total of $50 \times L$ composition differences were calculated, and the $50 \times L$ difference labels (positive, negative, no difference) were regarded as reference labels for performance evaluation. Labels positive and negative in the reference labels were considered as true positives ($s \neq 0$), with no difference labels considered as true negatives ($s = 0$).

For performance evaluation, we generated simulation data sets (S_1) consisting of 10 randomly selected immune cells and simulation data sets (S_2), consisting of 10 curationally selected immune cells. Then, we created three sets of replicates ($S_{1-1}, S_{1-2}, S_{1-3}$ and $S_{2-1}, S_{2-2}, S_{2-3}$), with different errors for each method.

Performance evaluation: To predict immune cell fractions by DCQ and ImmQuant, the expression matrix (E) of samples, where the proportion of immune cells is unknown such that cell surface marker gene expression data (matrix P), of known immune cells, are required as input Figure 3(a). To predict the immune cell fraction by LP, an expression matrix (E) of samples, where the proportion of immune cells is unknown, and cell surface marker gene expression of immune cells is known (reference table, R), are required as input Figure 3(b). We then created a reference table for 10 selected immune cells, and 25 cell markers. The reference table was created by integrating the markers of immune cells identified in previous studies (Novershtern et al., 2011), with those provided by BIO-RAD.

We then compared the Digital Cell Quantification (DCQ) (Altboum et al., 2014) algorithm, against our method, based on the simulation data sets (S_1, S_2). For this, we measured performances using the Receiver Operating Characteristic (ROC) curve, using the “pROC” R package (Robin et al., 2011).

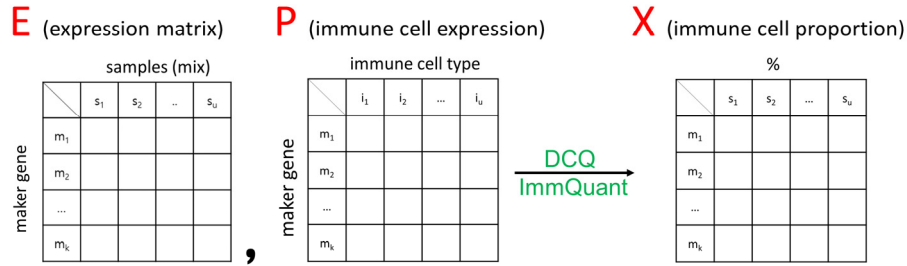
For immune cell composition prediction by DCQ, matrix P and a cell surface marker gene list were used as input, to predict immune cell composition, in matrix S . For this work, the DCQ algorithm used the “ComICS” R package (Altboum et al., 2014), with default options. From this analysis, we obtained the composition differences (i.e., the composition of sample class B minus that of the sample class A) of all the immune cell types, for each pair of classes A and B samples. These differences were further classified as either different ($s \neq 0$) or no difference ($s = 0$). We then inspected whether or not the differences predicted by DCQ agreed with the reference labels.

For immune cell composition prediction by the ImmQuant (Frishberg et al., 2016) tool (similar to DCQ), for which the matrix P and cell surface marker gene list were used as input, to predict immune cell compositions, in the matrix S . From such prediction, we obtained composition differences (i.e., the composition of the sample class B minus that of the sample class A), of all the immune cell types, for each pair of classes. The differences were then classified as either different ($s \neq 0$) or no difference ($s = 0$). Then, we

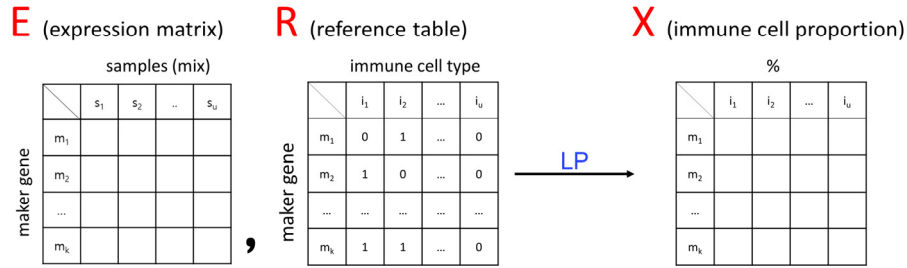
inspected whether or not the differences, predicted by the ImmQuant tool, agreed with the reference labels.

Figure 3 Comparing DCQ, ImmQuant and linear programming (LP). (a) DCQ and ImmQuant require, as input, an expression matrix (E), of unknown immune cell composition, and an expression matrix (P) of known immune cells. (b) LP requires, as input, an expression matrix (E) of unknown immune cell composition, and a reference table (R) of cell surface marker genes, of known immune cells

a.



b.



By our method, we also obtained immune cell compositions for S_A and S_B . The immune cell composition differences, for the 50 pairs between classes A and B , were labelled as either different ($s \neq 0$) or no difference ($s = 0$). We then compared our labels to the reference labels. In this task, we used the “fastclime” R package (Pang et al., 2014), with default options.

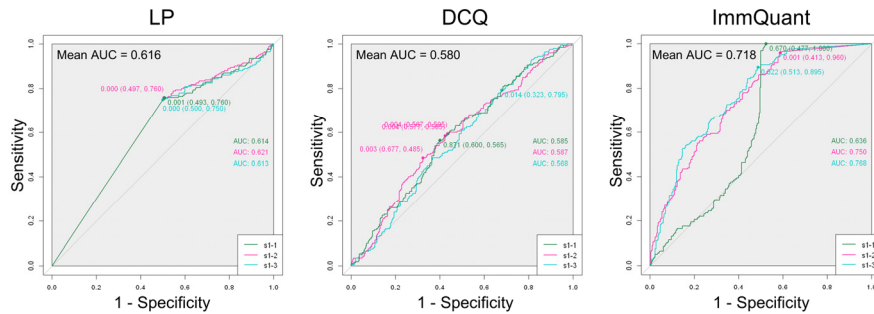
Systemic lupus erythematosus (SLE) patient data set: To apply this approach on a real biological data set, we imported two independent gene expression data (GEO accession: GSE49454 (Chiche et al., 2014), GSE61635) from the GEO (Barrett et al., 2013). We divided each data entry into SLE and healthy control groups. We selected 12 immune cell types from the Human Immune Cell Marker Tool®, provided by BIO-RAD, and used 24 markers to create a reference table, using the biomarker binary expression patterns in human immune cells data (“Biomarker Expression Patterns in Human Immune Cells” poster provided by BIO-RAD (Hercules, California, info.bio-rad.com)). We then predicted the fraction of immune cells, in each patient sample, using our LP method, and then compared the means of each group. We then confirmed the change in the fraction of immune cells in the SLE group, compared to the healthy control group.

3 Results

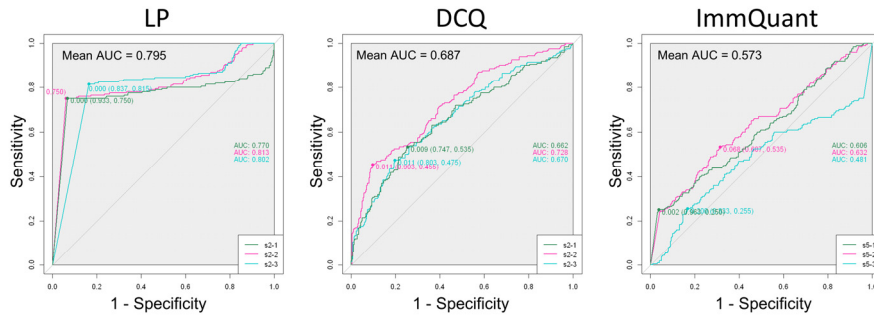
Simulated data set performance comparison: We first applied DCQ and our algorithms to a simulated data set, and drew ROC curves to compare Areas Under the Curve (AUCs) scores, to compare the performances of the two methods. In that regard, we measured their performances, using three duplicate data sets, and compared the means of the AUC values. Of the aforementioned in the methods section, 500 labels were generated by simulation. In simulation sets 1 (S_1), the mean of the AUC values was highest for the order of ImmQuant (0.718), LP (0.616), and DCQ (0.586) (see Figure 4(a)). In simulation set 2 (S_2), the mean of the AUC values was highest in the order of LP (0.795), DCQ (0.687), and ImmQuant (0.573) (Figure 4(b)). Comparing the performances, using two groups of simulation data sets, the AUC value range of the LP method was better than that of the AUC value range (0.616–0.795) of the ImmQuant (0.573–0.718) and DCQ (0.586–0.687) methods. Thus, from the variance of AUC values between subsets of each set, the difference was smaller for the LP method, compared to the other two methods.

Figure 4 Performance evaluation by simulation data. (a) Simulation dataset S_1 (randomly selected immune cells). Three subsets (S_{1-1} , S_{1-2} , S_{1-3}) were used to evaluate performance. The mean of the AUC values was highest in the order of ImmQuant (0.718), LP (0.616), and DCQ (0.586). (b) Simulation data set S_2 (curated selected immune cells). Three subsets (S_{2-1} , S_{2-2} , S_{2-3}) were used to evaluate performance. The mean of the AUC values was highest in the order of LP (0.795), DCQ (0.687), and ImmQuant (0.573)

a.



b.



Application of immune cell composition prediction in systemic lupus erythematosus (SLE) patients: We then compared changes in immune cell composition, between healthy controls and SLE patients. SLE, an autoimmune disease, is affected by both the innate and adaptive immune systems (Herrada et al., 2019), and it is straightforward to quantify immune cells as an approach for SLE biomarker discovery. So, we selected SLE as an application example of our LP method. In the first data set (GSE49454), our LP method reported that four of 12 immune cells were upregulated, and four cells downregulated, in SLE patients, compared to healthy controls (see Table 1). Previous SLE pathology studies, using the same data (Chiche et al., 2014), showed results for seven of 12 immune cells, and four of our cells were consistent with our results. In the second data set (GSE61635), our LP method reported that four of 12 immune cells were upregulated and four cells were downregulated, in SLE patients, compared to healthy controls (Table 1).

Table 1 Regulation of immune cell numbers. Regulation of 12 immune cell levels, in SLE patients, in two independent data sets, as predicted by the LP method

<i>Immune cell</i>	<i>GSE49454</i>	<i>GSE61635</i>
<i>B cell</i>	–	–
<i>Pro/Pre B cell</i>	Down-regulation	Up-regulation
<i>Myeloid dendritic cells</i>	Up-regulation	Up-regulation
<i>Neutrophil</i>	Up-regulation	Up-regulation
<i>Monocyte</i>	–	–
<i>Granulocyte</i>	Down-regulation	Down-regulation
<i>NK cells</i>	Down-regulation	Up-regulation
<i>T cells</i>	–	–
<i>Platelets</i>	Up-regulation	Down-regulation
<i>NKT cells</i>	Down-regulation	Down-regulation
<i>Megakaryocyte</i>	–	–
<i>Erythrocyte</i>	Up-regulation	Down-regulation

Note: – No change

4 Discussion

Using our gene expression data, we developed a model to study the fraction of immune cells in an uncharacterised sample. Unlike previous methods (Altboum et al., 2014; Frishberg et al., 2016), this was straightforward task, and a higher level of prediction was determined to be possible through Linear Programming (LP).

The regulation of immune cells did not exactly match between the predictions of the two data sets consisting of patients with Systemic Lupus Erythematosus (SLE). This may be due to the characteristics of the data set, and indicates the need to improve the LP method.

Our LP method has certain limitations, due to the nature of linear programming, such that the model cannot be extended to many cell types and markers. Here, we used only 24 of the 200 or more markers provided by BIO-RAD to predict the immune cell fraction from SLE data, and did not take into account the subsets of each immune cells. For

example, in the case of T cells, different subsets (e.g., CD8⁺ T cell vs. CD8⁻ T cell) exist, due to differentiation, and each subset has different cell markers (Golubovskaya and Wu, 2016).

However, it is important that we present a simple model for predicting the fraction of immune cells, using only the expression level of a small set of genes as markers. Although there is still room for improvement for immune cell quantification, using a small number of marker genes, our method may provide direction for future research, and possibly strategies for the therapy of disorders of the immune system.

Acknowledgements

This work was supported by the Gachon University research fund of 2018 (GCU-2018-0325). The study was also supported by the Gachon University Gil Medical Centre (grant number FRD2019-11 to S.N.).

References

- Altboum, Z., Steuerman, Y., David, E., Barnett-Itzhaki, Z., Valadarsky, L., Keren-Shaul, H., Meninger, T., Mendelson, E., Mandelboim, M., Gat-Viks, I. and Amit, I. (2014) 'Digital cell quantification identifies global immune cell dynamics during influenza infection', *Molecular Systems Biology*, Vol. 10. Doi: 10.1002/msb.134947.
- Bandura, D.R., Baranov, V.I., Ornatsky, O.I., Antonov, A., Kinach, R., Lou, X., Pavlov, S., Vorobiev, S., Dick, J.E. and Tanner, S.D. (2009) 'Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry', *Analytical Chemistry*, Vol. 81, No. 16, pp.6813–6822.
- Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C.L., Serova, N., Davis, S. and Soboleva, A. (2013) 'NCBI GEO: archive for functional genomics data sets – update', *Nucleic Acids Research*, Vol. 41, pp.D991–D995.
- Brodin, P. and Davis, M.M. (2017) 'Human immune system variation', *Nature Reviews Immunology*, Vol. 17, No. 1, pp.21–29.
- Chattopadhyay, P.K. and Roederer, M. (2015) 'A mine is a terrible thing to waste: high content, single cell technologies for comprehensive immune analysis', *American Journal of Transplantation*, Vol. 15, No. 5, pp.1155–1161.
- Chiche, L., Jourde-Chiche, N., Whalen, E., Presnell, S., Gersuk, V., Dang, K., Anguiano, E., Quinn, C., Burtsey, S., Berland, Y., Kaplanski, G., Harle, J.R., Pascual, V. and Chaussabel, D. (2014) 'Modular transcriptional repertoire analyses of adults with systemic lupus erythematosus reveal distinct type I and type II interferon signatures', *Arthritis and Rheumatology*, Vol. 66, No. 6, pp.1583–1595.
- Frishberg, A., Brodt, A., Steuerman, Y. and Gat-Viks, I. (2016) 'ImmQuant: a user-friendly tool for inferring immune cell-type composition from gene-expression data', *Bioinformatics*, Vol. 32, No. 24, pp.3842–3843.
- Golubovskaya, V. and Wu, L. (2016) 'Different subsets of t cells, memory, effect or functions, and CAR-T immunotherapy', *Cancers (Basel)*, Vol. 8, No. 3, pp.1–12.
- Han, A., Glanville, J., Hansmann, L. and Davis, M.M. (2014) 'Linking T-cell receptor sequence to functional phenotype at the single-cell level', *Nature Biotechnology*, Vol. 32, No. 7, pp.684–692.

- Herrada, A.A., Escobedo, N., Iruretagoyena, M., Valenzuela, R.A., Burgos, P.I., Cuitino, L. and Llanos, C. (2019) 'Innate immune cells' contribution to systemic lupus erythematosus', *Frontiers in Immunology*, Vol. 10. Doi: 10.3389/fimmu.2019.00772.
- Hobson-Gutierrez, S.A. and Carmona-Fontaine, C. (2018) 'The metabolic axis of macrophage and immune cell polarization', *Disease Models and Mechanisms*, Vol. 11, No. 8. Doi: 10.1242/dmm.
- Kaczorowski, K.J., Shekhar, K., Nkulikiyimfura, D., Dekker, C.L., Maecker, H., Davis, M.M., Chakraborty, A.K. and Brodin, P. (2017) 'Continuous immunotypes describe human immune variation and predict diverse responses', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 114, No. 30, pp.E6097–E6106.
- Kim, K., Yang, S., Ha, S.J. and Lee, I. (2020) 'VirtualCytometry: a webserver for evaluating immune cell differentiation using single-cell RNA sequencing data', *Bioinformatics*, Vol. 36, No. 2, pp.546–551.
- Lyons, Y.A., Wu, S.Y., Overwijk, W.W., Baggerly, K.A. and Sood, A.K. (2017) 'Immune cell profiling in cancer: molecular approaches to cell-specific identification', *NPJ Precision Oncology*, Vol. 1, No. 1. Doi: 10.1038/s41698-017-0031-0.
- Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., Trombetta, J.J., Weitz, D.A., Sanes, J.R., Shalek, A.K., Regev, A. and McCarroll, S.A. (2015) 'Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets', *Cell*, Vol. 161, No. 5, pp.1202–1214.
- Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., Frampton, G.M., Drake, A. C., Leskov, I., Nilsson, B., Preffer, F., Dombkowski, D., Evans, J.W., Liefeld, T., Smutko, J.S., Chen, J., Friedman, N., Young, R.A., Golub, T.R., Regev, A. and Ebert, B.L. (2011) 'Densely interconnected transcriptional circuits control cell states in human hematopoiesis', *Cell*, Vol. 144, No. 2, pp.296–309.
- Pang, H., Liu, H. and Vanderbei, R. (2014) 'The fastclime package for linear programming and large-scale precision matrix estimation in R', *Journal of Machine Learning Research*, Vol. 15, pp.489–493.
- Park, S. and Nam, S. (2019) 'sleBioRepo: a gene expression compendium in systemic lupus erythematosus', *International Journal of Data Mining and Bioinformatics*, Vol. 21, No. 4, pp.365–375.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C. and Muller, M. (2011) 'pROC: an open-source package for R and S+ to analyze and compare ROC curves', *BMC Bioinformatics*, Vol. 12, pp.1–9. Doi: 10.1186/1471-2105-12-77.
- Sturm, G., Finotello, F., Petitprez, F., Zhang, J.D., Baumbach, J., Fridman, W.H., List, M. and Aneichyk, T. (2019) 'Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology', *Bioinformatics*, Vol. 35, No. 14, pp.i436–i445.