Detecting Lineage-Specific Marker Genes for Tumor Evolution Based on Single Cell Transcriptome

Kosho Murayama, Hideo Matsuda*
Graduate School of Information Science and Technology, Osaka University, Suita, Osaka, Japan.

* Corresponding author. Tel.: +81-6-6879-4390; email: matsuda@ist.osaka-u.ac.jp
Manuscript submitted February 3, 2021; accepted April 10, 2021.
doi: 10.17706/ijbbb.2021.11.3.50-57

Abstract: Marker gene detection is important for understanding tumor progression; therefore, this paper proposes a novel approach to detect marker genes for each of several lineages as tumor cells progress. The approach clusters the tumor cells into groups that reflect the drug sensitivity, resistance, and other characteristics and infers lineages undergoing transitional processes along with tumor progression and potential branching. The approach estimates continuous gene expression during the progression of each lineage and identifies marker genes by analyzing within-lineage and between-lineage differential expression. Using this method, we successfully detected a potential marker gene related to tumor drug resistance, which was not detected by a conventional cluster-based method.

Key words: Marker genes, single-cell RNA-seq, trajectory inference, tumor progression.

1. Introduction

Single-cell RNA-seq (scRNA-seq) has been widely used for transcriptomic analysis at individual cell resolution. A substantial number of software tools have been developed [1], and each tool is classified into one or more categories according to its analytical task. One of the most common tasks is to assign cells to groups by ordering and clustering analyses. This task has been the largest area of development in single-cell analysis [1] and is often performed with clustering tools such as Seurat [2] for identifying the cell types in a sample. Then, trajectory analysis tools (for example, Monocle [3] and slingshot [4]) are used to reveal how gene expression changes across temporal processes such as development [5] and stimulation responses [6].

In this paper, we propose a novel approach to analyze the tumor evolution processes. The goal of our study was to identify the marker genes expressed in tumor cells during the transition from drug sensitivity to resistance. The current approaches for this purpose (for example, that reported by Sharma *et al.* [7]) often perform tumor cell clustering, identify drug-sensitive and drug-resistant groups of cells, and find differentially expressed genes between the two groups. However, these approaches do not consider the changes in gene expression in the middle of the transition process. Recently, Cook *et al.* analyzed temporal expression changes in genes in epithelial-mesenchymal transition (EMT) time-course experiments with lung, prostate, breast and ovarian cancer cell lines [8]. They assessed the gene expression dynamics throughout the pseudotemporal trajectories by using the R package psupertime [9]. To build the pseudotime model for each time course, psupertime requires sequential labels of the cells identified in the scRNA-seq data and identifies a linear combination of genes that places the cells in the specified label order.

Our approach consists of the following four steps: (1) clustering the tumor cells into groups reflecting

their drug sensitivity, resistance and other characteristics with the removal of any existing batch effects between the different tumor samples; (2) performing trajectory inference to identify lineages of transitional processes along with tumor progression and potential branching; (3) estimating the continuous gene expression of each lineage; and (4) identifying marker genes of both within-lineage and between-lineage differential expression. By using this approach, one can identify marker genes that are differentially expressed as specific lineages progress toward drug resistance.

2. Marker Analysis of Lineages of Tumor Evolution

2.1. Integration and Clustering of Single-Cell RNA-seq Data

We used scRNA-seq data to analyze tumor evolution from the Gene Expression Omnibus repository (GSE117872). The data were derived with scRNA-seq of cisplatin-sensitive, cisplatin-resistant, and cisplatin-resistant drug-holiday cells from patient-derived oral squamous cell carcinoma (OSCC) cells isolated from primary and metastatic sites [7]. The data included two primary OSCC patient-derived cell lines (HN120P and HN137P), synchronous (lymph node) metastatic cells (HN120M and HN137M), and drug-resistant cells (HN120PCR, HN137PCR, HN120MCR, and HN137MCR) generated with cisplatin treatment for four months. The data also included drug-holiday cells (HN120PCRDH, HN137PCRDH, and HN120MCRDH) that were treated with cisplatin and repopulated in the absence of cisplatin. We did not use the cell line HN148 because it contains only drug-sensitive cells but does not include drug-resistant and drug-holiday cells.

We integrated these HN120 and HN137 cell lines with the Seurat v3 integration method [2] and performed t-distributed stochastic neighbor embedding (t-SNE) dimensional reduction (Fig. 1(a)). We also conducted clustering with Seura t (resolution=1.0) (Fig. 1(b)).

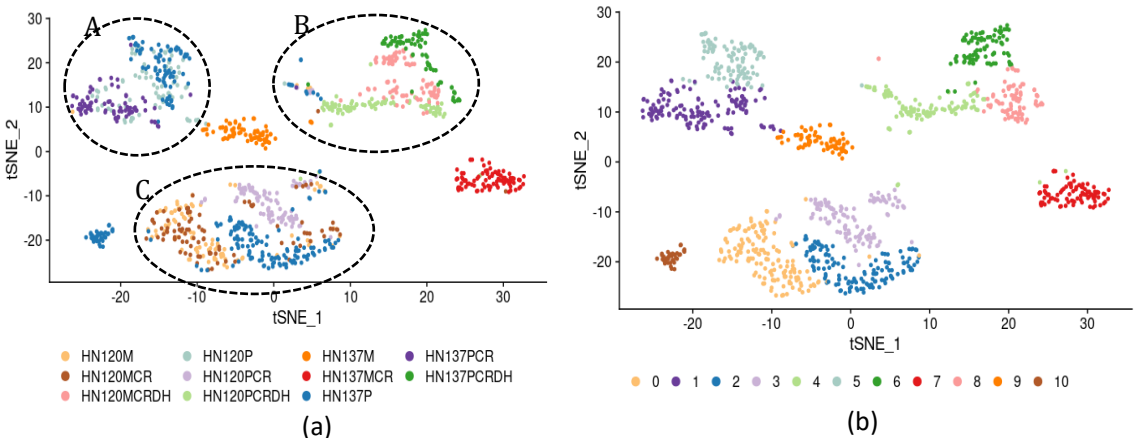

Fig. 1. t-SNE plots of tumor cells with cell types (a) and Seurat clusters (b).

As shown in Fig. 1(a), three large groups (A, B, and C) were detected. Groups A, B, and C consisted of HN120P, HN137P (partial), and HN137PCR; HN120PCRDH, HN120MCRDH, and HN137PCRDH; and HN120PCR, HN120M, HN120MCR, and HN137P (partial), respectively. HN137P (partial, lower left cluster), HN137M, and HN137MCR were not included in these three groups. Based on the cell types contained in the groups, we considered groups A, B, and C as primary (drug-sensitive), drug-holiday, and drug-resistant cells, respectively. A previous analysis [7] classified the data into five clusters (Fig. 2(d) in [7]). However, the clusters (presented in [7]) of drug-holiday cells (HN120PCRDH, HN120MCRDH, and HN137PCRDH) were separated into two clusters, while the integration method we used successively grouped the drug-holiday cells into the same group, group B.

In the clustering result, Fig. 1(b) shows that groups A, B, and C correspond to clusters 1 and 5; 4, 6, and 8; and 0, 2, and 3, respectively. HN137P cells were separated into three clusters (2, 5, 10). In [7], the cells also diverged into two clusters, and the authors [7] found that HN137P cells were heterogeneous populations. For this reason, in this paper, we removed the cells from cluster 10 for further analysis.

2.2. Lineage Analysis

We conducted trajectory inference with slingshot [4]. In the inference, we set cluster 5 (HN120P and HN137P (partial)) as a starting cluster and detected three lineages (Fig. 2). These lineages denote the lineage that progressed from drug-sensitive to drug-resistant (lineage 1), the drug-holiday lineage (lineage 2), and HN137MCR cells (metastatic and drug-resistant) (lineage 3). Interestingly, we identified a lineage that progressed from drug-sensitive to drug-holiday (lineage 2), although previous research [7] did not detect such a lineage.

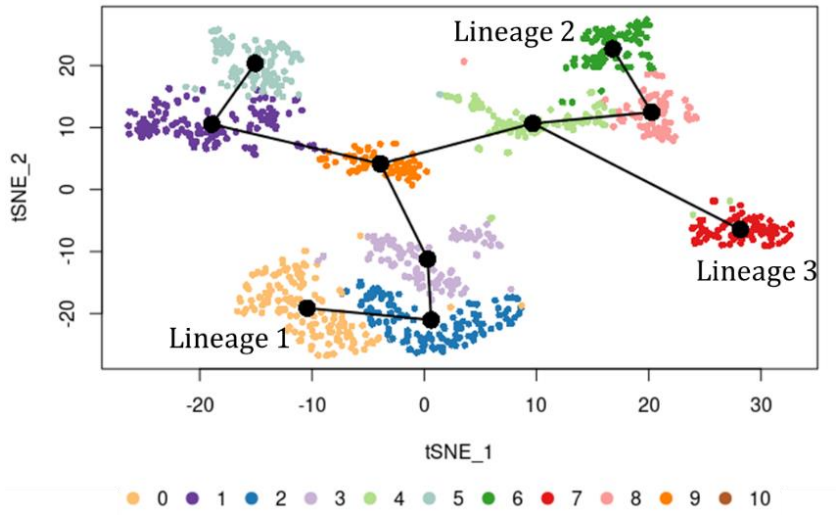

Fig. 2. Lineages detected by trajectory inference.

2.3. Estimation of Continuous Lineage-Specific Gene Expression

Given the lineages identified in Section 2.2, we estimated lineage-specific gene expression. To compute continuous gene expression across the clusters of the lineage, we employed tradeSeq [10] to infer smooth functions for gene expression along pseudotime for each lineage using generalized additive models.

To analyze lineage-specific changes between the drug-sensitive, drug-resistant, and drug-holiday states, we examined the expression of the marker genes for basal (KRT5), luminal (KRT18), epithelial (CDH1), and mesenchymal (VIM) cells (Fig. 3).

The smooth plots of marker gene expression make it possible to investigate the transition states of tumor cells in detail. As shown in Fig. 3, the cell-state transitions from basal/epithelial to luminal/mesenchymal were clearly detected by this analysis. Interestingly, the expression of a basal marker (KRT5) mostly decreases along pseudotime but increases only at the end of lineage 2 (drug-holiday lineage). This might suggest that the cells return to the primary (drug-sensitive) state in the absence of cisplatin. Additionally, a mesenchymal marker (VIM) was highly expressed at the pseudotime midpoint in lineage 3 (lineage toward HN137MCR). This may reflect that the transition to the mesenchymal state is related to metastasis and drug resistance.

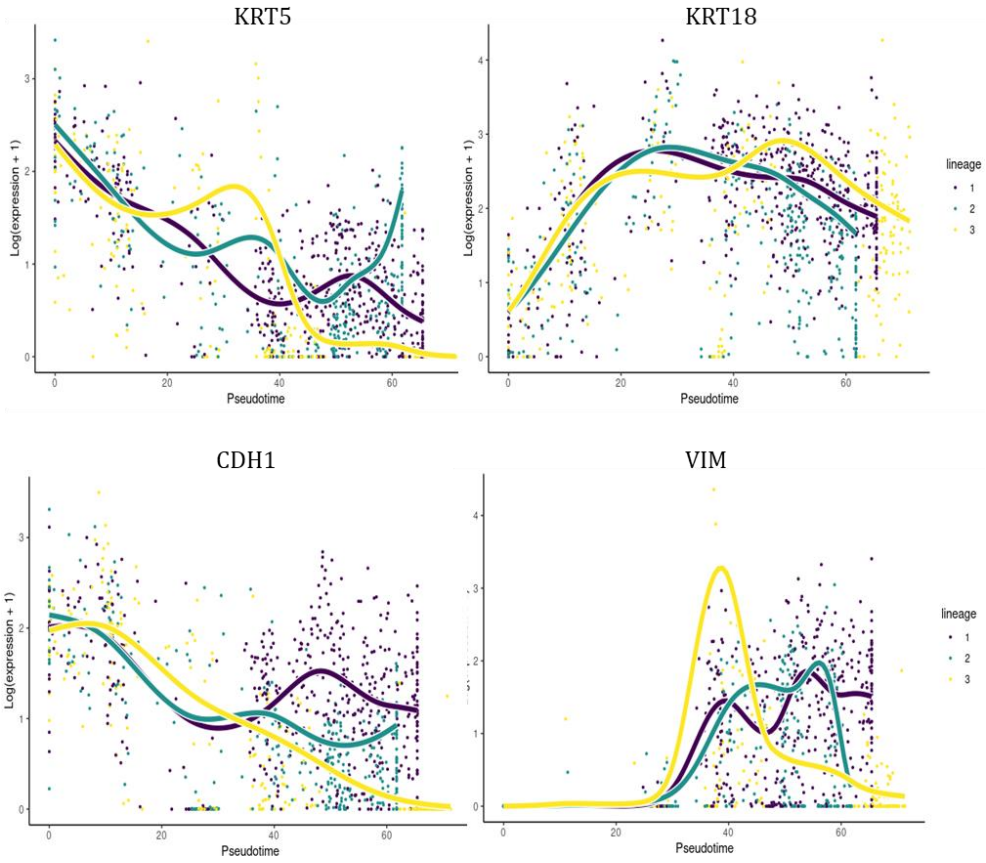

Fig. 3. Continuous changes in the expression of marker genes along lineage progression.

2.4. Identifying Marker Genes of within- and between-Lineage Differential Expression

We performed several statistical tests on expression within and between the lineages described in Section 2.2 by using tradeSeq [10]. To discover marker genes of the cell-state transitions of tumor progression, we assessed differential expression between the starting and ending points along lineages. We used a Wald test to assess the null hypothesis that the average expression at the starting point (drug-sensitive population) is equal to the average expression at the endpoints (drug-resistant, drug-holiday, drug-resistant with metastasis). Table 1 shows the marker genes of lineages 1, 2, and 3 detected by the test. Interestingly, the top marker gene detected in lineages 1 and 3 was the same, AXL. Fig. 4 shows the expression of the top marker genes.

Table 1. Marker Genes Detected by a within-Lineage Test

(a) Lineage 1 (Drug-Resistant)

	waldStat_lineage1	logFClineage1
AXL	143.44965	4.530163
SPC24	96.58933	3.944571
AAED1	83.95473	3.302275
UBE2C	74.66545	4.455476
NUSAP1	69.82497	3.438380

(b) Lineage 2 (Drug-Holiday)

	waldStat_lineage2	logFClineage2
PTGR1	60.14536	2.293527
ADA	54.45957	3.127771
UCA1	38.91781	3.843716
CCDC80	28.49954	5.182936
OLR1	24.46672	6.701984

(c) Lineage 3 (Drug-Resistant with Metastasis)

	waldStat_lineage3	logFClineage3
AXL	166.88115	4.645313
MT2A	83.50560	2.465299
HLA-B	78.05788	2.505961
UBE2C	72.25516	3.636388
HES1	67.30687	2.963632

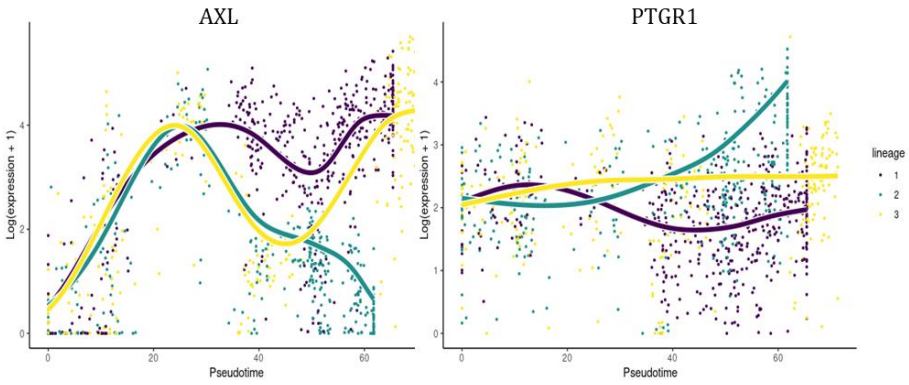

Fig. 4. Expression of the marker genes detected by the within-lineage test.

We also conducted a between-lineage test (lineages 1 and 3). We used a statistical method that determines whether the smoothed gene expression was equal between the two lineages along pseudotime. One hundred points, equally distributed along pseudotime, were compared between the two lineages. Table 2 shows the results, and Fig. 5 shows the expression of the top marker gene FTL.

Table 2. Marker Genes Detected by the between-Lineages Test

	waldStat_1vs3	fcMedian
FTL	520.4145	0.3783549
HES1	277.6275	0.7749512
F3	251.5828	0.4123971
LIMCH1	231.6139	0.9090507
CHP1	213.4125	0.4633066

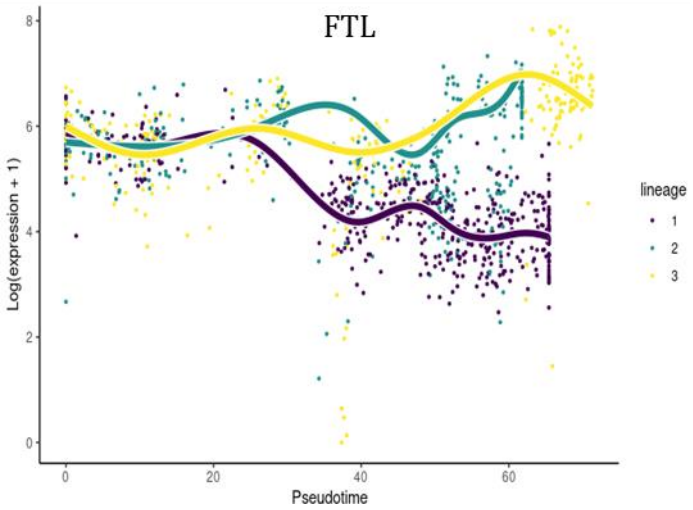

Fig. 5. Expression of a marker gene detected by the between-lineage test.

2.5. Comparison with Cluster-Based Marker Analysis

We compared the marker genes described in Section 2.4 to the cluster-based marker genes between groups A (drug-sensitive) and B (drug-resistant) by using Seurat v3 (FindMarkers with the options: only.pos=TRUE, min.pct=0.25). Interestingly, FTL was found (8th rank), but AXL was not included in the resulting list of cluster-based marker genes (data not shown). This suggests that our detection of marker genes along lineage progression can provide a different means for understanding tumor drug resistance mechanisms.

3. Discussion

It is key in cancer research to elucidate the resistance mechanism of cancer to the current standard of treatment. To date, many investigators have been working on developing an effective method to identify gene expression markers by going beyond bulk RNA-seq with scRNA-seq. In this paper, we conducted lineage-based differential expression analysis to identify marker genes that exhibit differential expression as specific lineages progress toward drug resistance.

As a result of the marker genes detected in the within-lineage test, we showed that the receptor tyrosine kinase AXL is a potential candidate to predict drug resistance under the selection pressure of cisplatin, which is one of the standards of treatment for OSCC. AXL is one of the TYRO3, AXL, MERTK (TAM) family receptor tyrosine kinases that plays an important role in promoting the growth, survival, and metastatic spread of several tumor types [11]. Based on previous research, AXL is known as a resistance mediator of cetuximab, which is an antibody targeting the epidermal growth factor receptor and one of the standards of treatment for head and neck cancer [12]-[14]. Therefore, based on our results, AXL could be involved in a potential resistance mechanism to cisplatin as well as cetuximab. In addition, as a result of the marker genes detected in the within-lineage test, we showed that the expression level of PTGR1 increased under the drug holiday. High expression of PTGR1 promotes cancer cell growth [15].

Furthermore, the marker genes detected by the between-lineage test that compared lineage 1 (drug-resistant) and lineage 3 (drug-resistant with metastasis) showed that FTL may be associated with resistance to metastasis. High expression of FTL could be considered a prognostic factor for several cancers, including glioma, colorectal cancer, ovarian cancer, gastric cancer, and acute myeloid cancer [16]-[20]. Moreover, high expression of FTL is considered to be involved in a chemotherapy resistance mechanism, including for 5-fluorouracil (5-FU) [17]. In addition, hypoxia-inducible FTL is a regulator of EMT [16].

4. Conclusion

In this paper, we propose a novel approach for the detection of marker genes as lineages transition between the drug-sensitive and drug-resistant states. Our approach can analyze continuous state transitions based on single-cell transcriptomic data, and it successfully detected a potential marker gene, AXL, which was not detected by a conventional cluster-based marker gene analysis tool. This lineage-specific marker gene analysis will provide an effective means to understand the processes of changes in cell state during tumor progression. Our future works include the development of tools that can identify more potential lineages as trajectory inference methods are improved.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Kosho Murayama and Hideo Matsuda conducted the research, analyzed the data, and wrote the paper; both authors approved the final version of the manuscript.

Acknowledgment

This work was supported in part by JST CREST Grant Number JPMJCR15G1; JSPS KAKENHI Grant Numbers JP18H04124, JP19K22894, and JP20H04947, Japan.

References

[1] Zappia, L., Phipson, B., & Oshlack, A. (2018). Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database, *PLOS Computational Biology, 14(6)*, e1006245.

[2] Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., *et al.* (2019). Comprehensive integration of single-cell data. *Cell, 177(7)*, 1888-1902.

[3] Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A., *et al.* (2017). Reversed graph embedding resolves complex single-cell trajectories, *Nature Methods, 14(10)*, 979-982.

[4] Street, K., Risso, D., Fletcher, R., Das, D., Ngai, J., Yosef, N., *et al*. (2018). Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics, 19(1)*, 477.

[5] Paul, F., Arkin, Y., Giladi, A., Jaitin, D. A., Kenigsberg, E., Keren-Shaul, H., *et al.* (2015). Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell, 163(7)*, 1663-1677.

[6] Osato, N., Shigeta, H., Seno, S., Uchida, Y., Kikuta, J., Ishii, M., *et al.* (2019). Single-cell transcriptome analysis of mouse leukocytes in inflammatory stimulation, *Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp.1229-1231). San Diego, USA.

[7] Sharma, A., Cao, E. Y., Kumar, V., Zhang, X., Leon, H. S., Wong, A. M. L., *et al.* (2018). Longitudinal single-cell RNA sequencing of patient-derived primary cells reveals drug-induced infidelity in stem cell hierarchy. *Nature Communications, 9(1)*, 4931.

[8] Cook, D. P., & Vanderhyden, B. C. (2020). Context specificity of the EMT transcriptional response. *Nature Communications, 11(1)*, 2142

[9] Macnair, W., & Claassen, M. (2019). Psupertime: Supervised pseudotime inference for single cell RNA-seq data with sequential labels. *Biorxiv Preprint*. Retrieved from https://doi.org/10.1101/622001

[10] Berge, K. V., Bézieux, H. R., Street, K., Saelens, W., Cannoodt, R., Saeys, Y., *et al.* (2020). Trajectory-based differential expression analysis for single-cell sequencing data. *Nature Communications, 11(1)*, 1201.

[11] Linger, R. M., Keating, A. K., Earp, H. S., & Graham, D. K. (2008). TAM receptor tyrosine kinases: Biologic functions, signaling, and potential therapeutic targeting in human cancer. *Advanced Cancer Research,*

100, 35-83.

[12] Brand, T. M., Iida, M., Stein, A. P., Corrigan, K. L., Braverman, C. M., Luthar, N., *et al.* (2014). AXL mediates resistance to cetuximab therapy. *Cancer Research, 74(18)*, 5152-5164.

[13] McDaniel, N. K., Iida, M., Nickel, K. P., Longhurst, C. A., Fischbach, S. R., *et al*. (2020). AXL mediates cetuximab and radiation resistance through tyrosine 821 and the c-ABL kinase pathway in head and neck cancer. *Clinical Cancer Research, 26(16)*, 4349-4359.

[14] Picon, H., & Guddati, A. K. (2020). Mechanisms of resistance in head and neck cancer. *American Journal of Cancer Research, 10(9)*, 2742-2751.

[15] Xue, L., Zhu, Z., Wang, Z., Li, H., Zhang, P., Wang, Z., *et al*. (2016). Knockdown of prostaglandin reductase 1 (PTGR1) suppresses prostate cancer cell proliferation by inducing cell cycle arrest and apoptosis. *Bioscience Trends, 10(2)*, 133-139.

[16] Liu, J., Gao, L., Zhan, N., Xu, P., Yang, J., Yuan, F., *et al*. (2020). Hypoxia induced ferritin light chain (FTL) promoted epithelia mesenchymal transition and chemoresistance of glioma. *Journal of Experimental and Clinical Cancer Research, 39(1)*, 137.

[17] Li, Z., Liu, J, Chen, H., Zhang, Y., Shi, H., Huang, L., *et al*. (2019). Ferritin light chain (FTL) competes with long noncoding RNA Linc00467 for miR-133b binding site to regulate chemoresistance and metastasis of colorectal cancer. *Carcinogenesis, 41(4)*, 467-477.

[18] Wang, L., Li, X., Mu, Y., Lu, C., Tang, S., Lu, K., *et al*. (2019). The iron chelator desferrioxamine synergizes with chemotherapy for cancer treatment. *Journal of Trace Elements in Medicine and Biology, 56*, 131-138.

[19] Zhang, L., Chen, Z., & Xu, A. (2017). FTL: A novel predictor in gastric cancer. *International Journal of Clinical and Experimental Pathology, 10(7)*, 7865-7872.

[20] Bertoli, S., Paubelle, E., Bérard, E., Saland, E., Thomas, X., Tavitian, S., *et al*. (2019). Ferritin heavy/light chain (FTH1/FTL) expression, serum ferritin levels, and their functional as well as prognostic roles in acute myeloid leukemia. *European Journal of Haematology, 102(2)*, 131-142.

Kosho Murayama received his B.Sc. degree in science from Kagoshima University, Japan in 2008, and he received his M.E. degree in medical science from Kyushu University, Japan in 2010. His research interests include genomic data analysis, gene regulatory networks, and gene expression analysis in the field of cancer.

Hideo Matsuda received his B.Sc. degree in physics from Kobe University, Japan in 1982, and he received his M.E. and Ph.D. degrees in computer science from Kobe University, Japan in 1984 and 1987, respectively. He has been a professor in the Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University since 2002. His research interests include genomic data analysis, gene regulatory networks, and gene expression analysis.