Visualizing Cancer Heterogeneity with Dynamic Flow

Teppei Nakano[†] and Kazuki Ikeda[⋆]

[†]Keio University School of Medicine, Tokyo 160-8582, Japan

keiohigh2nd@gmail.com

[⋆]Department of Physics, Osaka University, Toyonaka, Osaka 560-0043, Japan

kikeda@het.phys.sci.osaka-u.ac.jp

Abstract

We provide a new method for visualizing tumor heterogeneity. Intratumoral heterogeneity is a major obstacle to cancer treatment and a significant confounding factor in bulk-tumor profiling. While there are many researches trying to understand tumor heterogeneity by clustering RNA expression, little endeavor for generating perspicuous figures which enhance our intuitive comprehension of tumor heterogeneity has been made. In this article we propose a novel method for visualizing tumor heterogeneity using vector fields associated with cancer cells. This new approach endows a clear representation of tumor heterogeneity. Moreover in comparison with a simulation result, the panoptic figures clearly suggest mutation profiles depend on the global structure of a tumor, and the interaction among genes in different clusters are captured graphically.

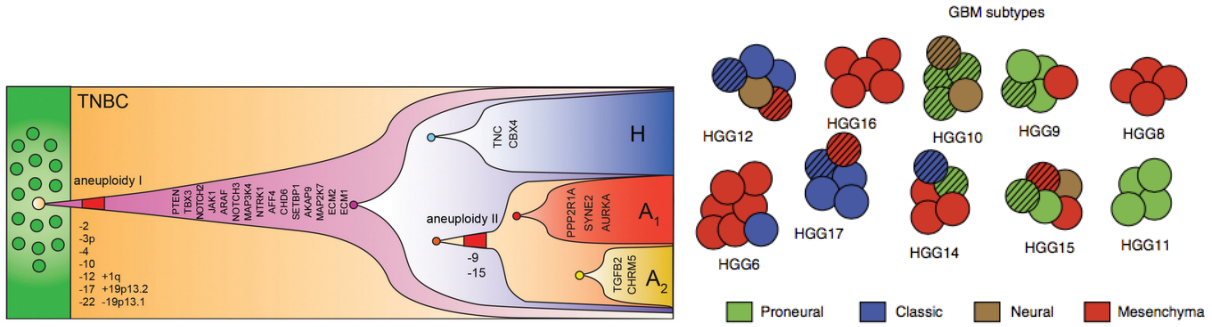
1 Introduction

Tumor heterogeneity describes the differences between tumors of the same type in various patients, and between cancer cells within a tumor [1, 2]. Detection of minor or generic distinct subpopulations within tumors is a cornerstone of cancer genomics [3, 4] since the subpopulations interact each other to survive from anti-cancer drugs and severe microenvironments. However it is difficult to understand the dynamics of the interactions among subpopulations, which makes it hard to utilize cancer heterogeneity in modern clinical situations.

Solving the heterogeneity of tumors attracts the interest of many researchers [5, 6, 7] from various viewpoints and one of the major approaches toward this issue is clustering bulk tumor into sub-groups. Those methods and visualizations give us sketchy details of the content and evolution of the bulk of tumors (Fig.1) [8, 9]. However they do not tell anything about correlations of cancer genes and cells in a tumor. It is paramount of importance to acquire detailed knowledge of the global structure of a tumor not only for providing efficient medical practice in clinical sites but also for designing ideal anticancer drugs, since the mechanism of revitalization of cancer genes is so complicated and kinetic that the conventional ways cannot capture the entire picture of the medical conditions [10].

In this work, we propose a novel method of visualizing tumor heterogeneity based on a well-defined mathematical framework using vector fields associated with cancer cells. This new approach enhances our comprehension of precise gene information of cancer tissues in a wide range of areas, which would lead to groundbreaking progress in developing antitumor agents. We also emphasize that simple machine learning techniques manifest power

greatly for this purpose. Especially, the figures we obtain in this way are in accord with simulation results of cancer evolution [11]. This work represents the first step for visualizing heterogeneity using machine learning.



(a) (b)

Figure1: Traditional methods to visualize heterogeneity. The tree structure in Fig. (a) shows wide variety of sub clonal population [9]. Proportion of cell types is represented in Fig. (b) with small dots [8].

2 Method

First of all, the methodological definition of tumor heterogeneity is not know, hence it is worth to consider what is a good definition of it. Conceptually, heterogeneity describes the difference of gene expression profiles. Therefor heterogeneity should be displayed by use of vector fields. We denote by $C = \{C_x\}$ a tumor which consists of cancer cells C_x, where C is assumed to be a smooth and connected two-dimensional manifold embedded in the three-dimensional Euclidean space $\mathbb{R}^3$, and the points of the cells are specified by local coordinates $x = (x, y)$ on C. Each of the cells has $N(= 300)$ genes and C_x is identified completely by measuring gene expression[1] $X_j(x)$[12]: $C_x = \{X_1(x), \cdots, X_N(x)\}$. We are interested in analyzing x-dependence of $X_j(x)$, hence we consider the space $\{x, dX_1(x), \cdots, dX_N(x)\}$, where $dX_j(x) = \frac{\partial X_j(x)}{\partial x}dx + \frac{\partial X_j(x)}{\partial y}dy$ $(j = 1, \cdots, N)$. The operation for generating an anticipated figure associated with the vector fields is performed by the following algorithm.

Algorithm

1. Based on RNA expression data, we list up genes that are statistically up regulated or down regulated compared to normal tissues.
2. Extracting the top up to 300 genes by absolute fold-change.
3. Compare the expression between combination of tumours cells. Calculate gradient using above.

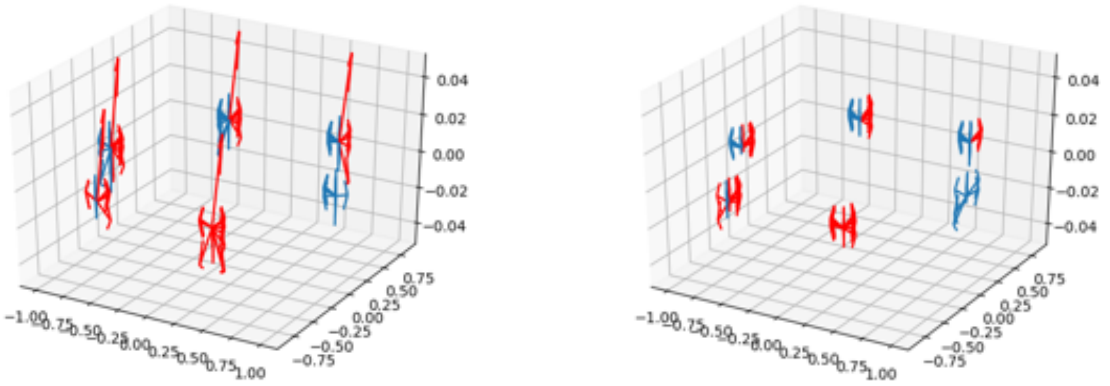
[1]We set $X_1 = \cdots = X_N = 0$ for normal tissues.

3 Experiment, Results and Analysis

Here we mention our main results which clearly show the nontrivial global structure of a tumor. We borrowed the experimental data from [3]. We used 4 High-grade glioma samples and 3 primary medulloblastoma samples. Each tumor was cut into 6 parts. RNA was extracted and hybridized to GeneChip® Human Genome U133 Plus 2.0 Array according to the manufacturer's instructions. Microarray extracted 54677 expression of genes. We compared gene expression profiles between normal and cancer tissues using EdgeR.

For simplicity, we set the marked point at x_1 and we replace $dX_j(x)$ by $X_j(x_i) - X_j(x_1)$ since cancer cells in the data set are isolated, where we denote by x_i $(i = 1, \cdots, 6)$ the points of the cells. Following the procedure proposed in the previous section, figures showing heterogeneity are obtained (Fig.2). The left figure (α) indicates high heterogeneity tumor bulks and the right (β) is about low heterogeneity tumor bulks. The x-axis and the y-axis label the local position of each sub populations. The z-axis stands for fold-change of gene expression. At each site x_i $(i = 1, \cdots, 6)$, arrows associated with the gene expression are shown and they are colored by the following rule:

$$dX_j \geq 0 \Leftrightarrow \text{RED}$$
$$dX_j < 0 \Leftrightarrow \text{BLUE}.$$



(α) High heterogeneity (β) Low heterogeneity

Figure 2: Heterogeneity of a tumor. For the presentational convenience, we took 60 genes expression profiles.

These figures clearly prove that the global structure within a tumor is extremely nontrivial and quite complex. At first sight, both of them tell us that the ratio of gene expression depends on the site of cells. Moreover they suggest there are gene expression patterns

which closely related with positions of the cells. Furthermore comparing (α) and (β), we find that fold-change of gene expression profiles in (β) is bigger than that in (α), which clearly explains that (α) shows high heterogeneity and (β) shows low heterogeneity. Our results suggest that using the vector field expression is a good way to capture the conceptual definition of tumor heterogeneity. The significance of the above figures can be explained as follows. The existence of a set of cells which have various combinations of gene expression, which we call heterogeneity, is a reason for cancer being treatment-resistant. For instance, high heterogeneity shows resistance to many anticancer drugs. Suppose there are cells of high heterogeneity (α) and cells of low heterogeneity (β) within a tumor, then the number of cells (β) becomes smaller as the treatment progress, while increase in the number of cells (α) triggers off redeveloping. Therefore revealing the heterogeneity is clinically crucial.

Appearance of heterogeneity is demonstrated by use of a supercomputer [11], and in comparison with the simulation results (the figures below), Fig. 2 captures the feature accurately. In the figure (A), differently colored cell populations represent each clone. Moreover,

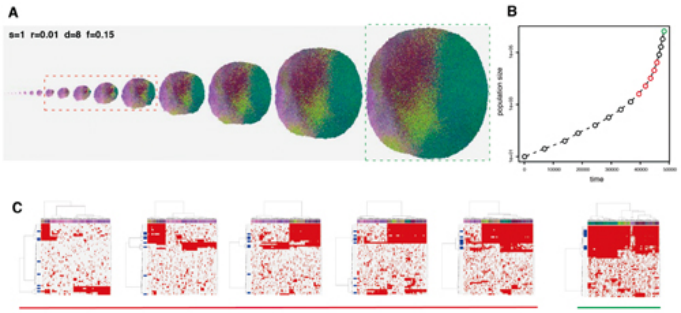


Figure 3: (A) shows growing tumors in a simulation and (B) represents its growth curve. Gene mutation patterns during growth are given in (C).

gene mutation profiles (C) show tumor heterogeneity. Consequently those figures $(\alpha), \cdots, (C)$ strongly suggest that gene expression patterns are totally position dependent and ruled by some mechanism.

4 Conclusion and Further Directions

Here we summarize our results and proposals. In a methodological way, we firstly defined tumor heterogeneity by use of vector fields associated with gene expression profiles. Secondly, we visualized tumor heterogeneity with this concept and confirmed that heterogeneity in the given cancer data we used would be easily understood. As we discussed, it is worth to mention that visualizing the cancer heterogeneity is tremendously important to understand the global structure of a tumor, while past studies focused on the local content of cancer heterogeneity. Moreover our approach is the first step in this direction and more sophisticated ways will helpful to uncover a mechanism of evolution of cancer cells. We also suggest it is necessary to collect more datas from many points in order to improve accuracy of the analysis and uncover more precise structure of a tumor.

References

[1] I. J. Fidler and I. R. Hart, "Biological diversity in metastatic neoplasms: origins and implications," *Science* **217** no. 4564, (1982) 998–1003.

[2] D. W. Parsons *et al.*, "An integrated genomic analysis of human glioblastoma multiforme," *Science* **321** no. 5897, (2008) 1807–1812.

[3] A. P. Patel *et al.*, "Single-cell rna-seq highlights intratumoral heterogeneity in primary glioblastoma," *Science* **344** no. 6190, (2014) 1396–1401.

[4] A. Marusyk, V. Almendro, and K. Polyak, "Intra-tumour heterogeneity: a looking glass for cancer?," *Nat Rev Cancer* **12** no. 5, (05, 2012) 323–334.

[5] N. E. Navin, "The first five years of single-cell cancer genomics and beyond," *Genome research* **25** no. 10, (2015) 1499–1507.

[6] M. Ladanyi and W. Pao, "Lung adenocarcinoma: guiding egfr-targeted therapy and beyond," *Mod Pathol* **21** no. S2, (Print, 0000) S16–S22.

[7] L. V. Sequist and T. J. Lynch, "Egfr tyrosine kinase inhibitors in lung cancer: an evolving story," *Annu. Rev. Med.* **59** (2008) 429–442.

[8] F. R. Hirsch, G. Ottesen, J. Pødenphant, and J. Olsen, "Tumor heterogeneity in lung cancer based on light microscopic features," *Virchows Archiv A* **402** no. 2, (1983) 147–153.

[9] Y. Wang *et al.*, "Clonal evolution in breast cancer revealed by single nucleus genome sequencing," *Nature* **512** no. 7513, (08, 2014) 155–160.

[10] S. Yachida *et al.*, "Distant metastasis occurs late during the genetic evolution of pancreatic cancer," *Nature* **467** no. 7319, (2010) 1114–1117.

[11] A. Niida, S. Ito, G. Tremmel, S. Imoto, R. Uchi, Y. Takahashi, Y. Takahashi, K. Mimori, and S. Miyano, "Cancer evolution simulation identifies possible principles underlying intratumor heterogeneity," *bioRxiv* (2015) .

[12] M. Kanamori-Katayama *et al.*, "Unamplified cap analysis of gene expression on a single-molecule sequencer," *Genome research* **21** no. 7, (2011) 1150–1159.

[13] L. True *et al.*, "A molecular correlate to the gleason grading system for prostate adenocarcinoma," *Proceedings of the National Academy of Sciences* **103** no. 29, (2006) 10991–10996.

[14] D. F. Gleason and G. T. Mellinger, "Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging.," *The Journal of urology* **111** no. 1, (1974) 58–64.