Seminars in Cancer Biology 23 (2013) 279–285

Contents lists available at SciVerse ScienceDirect

Seminars in Cancer Biology

journal homepage: www.elsevier.com/locate/semcancer

Review

Cancer systems biology in the genome sequencing era: Part 1, dissecting and modeling of tumor clones and their networks

Edwin Wang[a,b,*], Jinfeng Zou[a,c,d], Naif Zaman[a,e], Lenore K. Beitel[c,d], Mark Trifiro[c,d], Miltiadis Paliouras[c,d]

[a] National Research Council Canada, Montreal, Canada
[b] Center for Bioinformatics, McGill University, Montreal, Canada
[c] Lady Davis Institute, Montreal, Canada
[d] Department of Medicine, McGill University, Montreal, Canada
[e] Department of Anatomy and Cell Biology, McGill University, Montreal, Canada

ARTICLE INFO

Keywords:
Genome sequencing
Systems biology
Molecular network
Signaling network
Cancer hallmark
Tumor clonal network

ABSTRACT

Recent tumor genome sequencing confirmed that one tumor often consists of multiple cell subpopulations (clones) which bear different, but related, genetic profiles such as mutation and copy number variation profiles. Thus far, one tumor has been viewed as a whole entity in cancer functional studies. With the advances of genome sequencing and computational analysis, we are able to quantify and computationally dissect clones from tumors, and then conduct clone-based analysis. Emerging technologies such as single-cell genome sequencing and RNA-Seq could profile tumor clones. Thus, we should reconsider how to conduct cancer systems biology studies in the genome sequencing era. We will outline new directions for conducting cancer systems biology by considering that genome sequencing technology can be used for dissecting, quantifying and genetically characterizing clones from tumors. Topics discussed in Part 1 of this review include computationally quantifying of tumor subpopulations; clone-based network modeling, cancer hallmark-based networks and their high-order rewiring principles and the principles of cell survival networks of fast-growing clones.

1. Introduction

Advances in high-throughput technologies have made a great impact on revolutionizing medical research. Biology is becoming a data-intensive science through increasing use of "-omics" technologies. For example, whole-genome sequencing (WGS) has the potential to serve as a powerful and cost-effective diagnostic tool in the management of cancer. Tumor genome sequencing using WGS or WES (whole exome sequencing, i.e., sequencing of the coding regions of the genome) has provided huge amount of data for cataloging genomic alterations of tumors including copy number variations (CNVs), insertions, deletions, and single-nucleotide variants (SNVs). It is expected that 10,000 tumor genomes will be completed using WES or WGS by the end of 2014. Today, sequencing a genome takes 1–2 days and costs ~$5000. This means that genome analysis is now in the cost range of a sophisticated medical test such as magnetic resonance imaging. As sequencing technologies mature and costs are lowered, there has been an increase in the application of these technologies to tumor management.

Cancer is driven by changes in the genomes of cancer cells. Tumor genome sequencing efforts allow comprehensive cataloging of genomic alterations. Thus far, these catalogs have shown that the key mutated genes and pathways that are altered in cancer are already well-known. EGFR, RAS, PI3K, P53, FGFR, MET and many other well-known cancer-driver mutating genes have been frequently rediscovered in many tumor genome sequencing studies [1–3]. Similarly, previously known key cancer-driving pathways such as RAS-pathway, PI3K-pathway, EGFR-pathway, MAPK-pathway and so on have been documented again and again in sequencing many types of tumors [1–3]. These results basically confirmed the major cancer biology knowledge generated via previous small-scale studies. The cataloging activity does not provide many novel insights into the fundamental biology of cancer, and the catalogs alone may not unveil new cancer treatment strategies for cancer management or cure. However, the tumor sequencing activity does provide multiple types of data for tumors: mutation, CNVs, epigenetic and gene expression profiles.

Traditional cancer biology studies have focused on one gene or one pathway at a time. Given such complex data sets derived from tumor genome sequencing, it is possible to gain new insights by

* Corresponding author at: 6100 Royal Mount Avenue, Montreal, QC, Canada H4P 2R2. Tel.: +1 514 496 0914; fax: +1 514 496 5143.
E-mail address: edwin.wang@cnrc-nrc.gc.ca (E. Wang).

1044-579X/$ – see front matter.
http://dx.doi.org/10.1016/j.semcancer.2013.06.002

taking an integrative systems biology approach, instead of simply cataloging genomic alterations. For example, a systems approach tends to develop new methods for integrating genomic alterations, functional information on genes and molecular networks to model cancer development, metastasis, and drug resistance at both a personal and systems level to identify new treatment strategies [4,5]. Although tumor genome sequencing provides few new insights into cancer biology, it allows dissection of subpopulations (clones) of tumors and generation of insights into tumor evolution. These results have profound impact on cancer systems biology and led researchers to reconsider how to conduct cancer systems biology studies. In this review, we will highlight our thinking about new directions for systems biology studies of cancer in the context of genome sequencing.

2. Quantifying tumor subpopulations via genome sequencing

In 1976, Novell [6] proposed that a single cell could randomly acquire a series of mutations that allow it to proliferate, then differently mutated cells, or clones could compete with each other, and one clone could outcompete others and finally form a tumor. Novell's hypothesis emphasized that only one clone (i.e., the one finally outcompetes others) was the most fit to survive and the other less-fit clones die out within a tumor. In late 80s, studies reported that a linear accumulation of specific genetic changes convert a normal epithelial cell into a tumor, in which a single clone is dominant [7–9]. These results strongly support Novell's hypothesis. However, later clinical treatments hint that a tumor is heterogeneous, i.e., a tumor could contain more than one clone. Recent tumor genome sequencing studies [10,11] confirmed that many distinct subpopulations of cells, or clones, co-exist in a tumor. Genome sequencing reveals the genetic record of their emergence over time and allows us to trace the divergence of a cell to form the different clones. By the time the cancer is diagnosed, one of these clones has become the dominant population in the tumor.

For quantifying subpopulation of tumors, new computational algorithms such as ASCAT [12] and ABSOLUTE [13] have been developed. These tools are able to: (1) infer the number of clones and their fractions within a tumor; (2) generate genetic scripts or profiles of genomic alterations including mutations and CNVs for each clone; and (3) infer the order of occurrence (timing) of the clones or the evolutionary tree of the clones. These tools opened a new window into the complexity of cancer, through reconstructing genetic networks of genomic alterations for clones of a patient's tumor, and then studying individual clones at a systems level. Some clones (early-occurring clones) occur in early stages of tumorigenesis, while some (late-occurring clones) appear in late stages. Based on selective growth advantages of the clones, we classify the clones into slow-growing and fast-growing clones. Slow-growing clones cannot form clinically detectable tumors on their own, whereas fast-growing clones are able to grow sufficiently to be clinically detectable as tumors without extra genomic alterations. In general, early-occurring clones are slow-growing clones, while late-occurring clones are fast-growing clones. Based on the studies of tumor genome sequencing, we proposed three genetic models of tumorigenesis using breast cancer as an example (Fig. 1): (1) *the chromothripsis-driven model* [14,15]. Chromothripsis describes a process in which one or several chromosomes are shattered into hundreds of fragments in a single cellular catastrophe, and then the DNA repair machinery pastes them back together in a highly erroneous order. This process will generate gene amplifications/deletions on a massive scale. This is a fast track to form tumors which could contain very few clones. Only 2–5% of tumors may be generated in this fashion; (2) *the gradual mutation model*. This model suggests that cells accumulate mutations gradually and continuously to form slowing-growing clones first and then form fast-growing clones. The transition from slow-growing clones to fast-growing clones could be triggered by a genome duplication event (see Part 2 of this review [16]). One clone could have several direct daughter clones, thus, the clone populations form a family tree; and (3) the *stem cell model* [15]. This is similar to the gradual mutation model, but a stem cell clone could be formed at an early stage. A cancer stem cell clone generates new daughter cells, but with a limited and slow growth rate due to the unfavorable conditions in the microenvironment. These daughter cells further acquire new genomic alterations and then generate new clones. Clone populations in the tumor are also organized as a family tree, however, along the tree there might have a clone (stem cell clone) which could have significantly more branches than others.

Tumors formed via the gradual mutation or stem cell models could have one dominant fast-growing clone (e.g., representing 80–90% of a tumor volume), but most of the tumors could have a few fast-growing clones (the dominant one could represent 40–50% of a tumor volume). With these new understandings, and equipped with tools for dissecting and quantifying tumor clones, we should reconsider the strategies for conducting cancer systems biology studies, understanding the genetic/epigenetic underpinnings of human cancer, developing new insights into how to tackle this terrible disease, and finally new, more personalized treatments for cancer patients.

3. From tumor-based network modeling to clone-based network modeling

In the past, we studied a tumor as a whole entity. For example, we have generated enormous amounts of "omic" data for tumor samples in the past two decades. These data include gene microarray, RNA-seq, SNP, CNV and epigenetic profiles for all kinds of tumors. About ten years ago, a systems biology approach was applied to these data by focusing on modeling of molecular networks. However, the network approach has also considered a tumor as a single entity. In fact, all these data are a mixture of the profiles of multiple clones. Without dissecting clones, it is a black box for the number of clones and the fractions of each clone within a tumor. A tumor-based network thus describes 'crude' or inaccurate molecular interactions/regulatory relationships within the cancer cell. Therefore, tumor-based network modeling efforts have captured a large amount of noise, although some strong signals could be uncovered [17,18]. Without quantifying tumor clones, network modeling missed many important features of tumors, such as heterogeneity, which could be derived from clonal backup (i.e., within a tumor, when a clone is killed by a drug, another clone could overtake it and make the tumor recur). However, at present, clinically, a drug only targets one clone within a tumor. Thus, tumor-based network modeling is ineffective in both uncovering fundamental insights into cancer biology and providing clues for better treatment of cancer patients.

Even knowing these drawbacks, in the past we have had to model cancer cells using tumor-based networks due to the lack of means of dissecting clones from tumors. Today, tumor genome sequencing provides an opportunity to model cancer cells by constructing and modeling networks for individual clones. By doing so, we could overcome the drawbacks of tumor-based network modeling and provide more accurate and comprehensive understanding of tumors, for example, we could model not only each clone's network, but also the potential functional interactions between clones within a tumor. Mostly importantly, clone-based network modeling could capture the features of individual clones (e.g., some clones have aggressive/invasive features, while other clones have no