

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

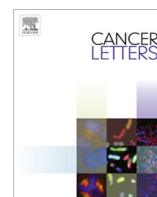
Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/authorsrights>

Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

Cancer Letters

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

Mini-review

Understanding genomic alterations in cancer genomes using an integrative network approach



Edwin Wang

Lab of Bioinformatics and Systems Biology, National Research Council Canada, Montreal, Canada
 McGill University Center for Bioinformatics, Montreal, Canada

ARTICLE INFO

Keywords:

Chromothripsis
 Kataegis
 Cancer genome evolution
 Network
 Systems biology
 Cancer genome sequencing

ABSTRACT

In recent years, cancer genome sequencing and other high-throughput studies of cancer genomes have generated many notable discoveries. In this review, novel genomic alteration mechanisms, such as chromothripsis (chromosomal crisis) and kataegis (mutation storms), and their implications for cancer are discussed. Genomic alterations spur cancer genome evolution. Thus, the relationship between cancer clonal evolution and cancer stem cells is commented. The key question in cancer biology concerns how these genomic alterations support cancer development and metastasis in the context of biological functioning. Thus far, efforts such as pathway analysis have improved the understanding of the functional contributions of genetic mutations and DNA copy number variations to cancer development, progression and metastasis. However, the known pathways correspond to a small fraction, plausibly 5–10%, of somatic mutations and genes with an altered copy number. To develop a comprehensive understanding of the function of these genomic alterations in cancer, an integrative network framework is proposed and discussed. Finally, the challenges and the directions of studying cancer omic data using an integrative network approach are commented.

Crown Copyright © 2012 Published by Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Because of the ongoing development of new, fast, and inexpensive DNA sequencing technologies, the cost of genome sequencing technology has rapidly declined. Eventually, genome sequencing technology may allow doctors to decoding the entire genetic code of a patient disease sample in a clinical setting. Cancer is rooted in genetic and epigenetic alterations, which are either inherited or acquired (i.e., mutated or methylated) during our lives. Cancer genomes contain two classes of mutations: driver mutations, which are positively selected because they are essential for tumor growth and development, and passenger mutations, which are not subject to selection because they do not offer a growth advantage. Positive selection indicative of driver mutations is evidenced by a higher probability than by chance of amino acid-changing nonsynonymous mutations to synonymous mutations that do not involve amino acid changes.

However, we are still no closer to uncovering the molecular underpinnings of the disease. If we could catalog cancer driver-mutating genes, we could link these genes to biological pathways, biological processes and cellular networks. Cancer genome sequencing and omic studies will improve the understanding of the genetic/epigenetic underpinnings of major types and subtypes of human cancer, thus generating new insights concerning how to

address and to manage cancer in clinics. These studies will eventually enable more personalized treatments that will improve patient care.

Significant effort has focused on compressive omic analysis, including genome sequencing, SNPs profiling, gene expression, epigenetic changes and proteomic profiling of major tumor types. These efforts represent a decade-long program that has aimed to generate high-quality omic data on more than 25,000 tumors for up to 50 types of cancer of clinical and societal importance. Three major organizations, the Wellcome Trust Sanger Institute in the United Kingdom, the Cancer Genome Atlas (TCGA) in the United States and the International Cancer Genome Consortium (ICGC) [1] in Canada, have advanced this program. Furthermore, several countries are working on this project, and the ICGC is collecting, merging and managing all of the data. In fact, the sequencing of 25,000 tumors is being performed around the world in 40 projects focused on the bladder, the blood, the bone, the brain, the breast, the cervix, the colon, the head and the neck, the kidney, the liver, the lung, the oral cavity, the ovary, the pancreas, the prostate, the rectum, the skin, the soft tissues, the stomach, and the uterus. The ICGC has collected omic data on more than 3500 tumors.

Among these efforts, St. Jude Children's Research Hospital took an initiative to make cancer genome information freely accessible to the global scientific community at www.ebi.ac.uk/ega/organisations/EGAO00000000046 [2]. This site mainly includes the genome sequences (entire genome sequencing with up to a 30-fold cover-

E-mail address: edwin.wang@cnrc-nrc.gc.ca

age) for matched sets of normal and tumor tissue samples from pediatric cancer patients. It is expected to sequence more than 1200 genomes by the end of 2012. It represents the world's greatest effort and investment to understand the genetic origins of childhood cancers and is the first major privately funded human genome sequencing project to share its data as soon as they become available. Table 1 summarizes the major completed genome sequencing efforts (i.e., at least 20 genomes have been sequenced in each study), most of which were published in 2012. In terms of cancer genome sequencing, 2012 signals the start of a significant increase in the amount of cancer genome sequencing data.

These efforts have generated several notable insights, including new genomic alteration mechanisms, the tracing and deciphering of cancer genome evolution and the dissection of tumor subclones.

2. Cancer genome sequencing uncovered unexpected genomic alteration mechanisms

2.1. Chromothripsis

Traditionally, cancer has been believed to develop gradually in a steady and stepwise progression. It often takes years to accumulate the multiple mutations to facilitate cancer's aggressive growth. In fact, most cancers undergo phases of abnormal tissue growth before developing into malignant tumors. However, recent analyses of tumor genomes revealed different patterns of genomic alteration mechanisms. One mechanism, "chromothripsis" [3], is a process by which a genome shatters into hundreds of fragments for one chromosome or several chromosomes in a single cellular catastrophe, and the DNA repair machinery then recombines the

fragments in a highly erroneous order. This process generates gene amplifications/deletions on a massive scale. This mechanism suggests that a chromothripsis (also called a chromosomal crisis) can generate multiple oncogenes or tumor suppressors in a single event. Chromosomal crises have been observed across all common cancers at an incidence rate of 1/40 but are more common in bone cancers, with an incidence rate of 20% [3]. Chromothripsis have been observed in other studies, although at a similarly low incidence rate (2–4%) [4,5].

It has been suggested that P53 mutation may trigger a chromothripsis. However, a P53 mutation is very common in cancer. An intriguing question concerns whether a chromosomal crisis occurs often or occasionally in the human body's cells. Notably, some cells that undergo a chromosomal crisis are able not only to survive but also to emerge with a genomic landscape that grants a selective advantage to the clone, thus promoting cancer development. Because a chromosomal crisis induces many dramatic genomic changes, we expect that a fraction of the genetic alterations are harmful to cells' survival. Should this expectation be accurate, many cells undergoing a chromosomal crisis could die, and we would be unaware of the occurrence. Another possibility is that tumor cells with many dramatic genetic alterations induced by a chromosomal crisis are so weak that they can be eliminated by tumor microenvironments. These factors may explain the extremely low rate of chromothripsis in tumors. Comparing the microenvironments of the tumors that have or do not have chromothripsis might offer insight regarding this point. The survival outcomes of the patients with chromothripsis-derived and patients with non-chromothripsis-derived tumors need to be reviewed.

2.2. Kataegis

Chromothripsis refers to a new oncogenic mechanism that operates on a global level; that is, it occurs in one chromosome or in several chromosomes. Another new mutation mechanism has been found to operate locally, generating large numbers of mutations (or hotspots of hypermutations) in small regions of the genome; this process has been termed 'kataegis' (from the Greek word for thunderstorm) [6]. Compared with chromothripsis, which occurs at a rate of 1/40 in cancer genomes, kataegis is remarkably common, occurring at a rate of 13/21 in breast cancer genomes. Kataegis most likely causes mutations to occur at one time rather than accumulating in a step-wise fashion. Kataegis is observed when a large number of substitution mutations occur simultaneously and extremely close to one another in a region of the genome [6]. Therefore, kataegis can be identified as distinguishing mutational patterns that often co-occur with large-scale rearrangements. Kataegis has been described as most likely occurring at one time; however, certain regions of the genome may mutate repeatedly during tumor evolution. If this possibility is proven, the affected regions are extended due to mutations between different subclones. Whether such extended mutation profiles are found in these genome regions requires further examination.

These mutation mechanisms indicate that internal factors, such as genome rearrangement, may contribute some types of genetic mutations and that some regions of cancer genomes are more likely to mutate than others. Several questions remain. The first question concerns (1) whether kataegis occurs once or repeatedly in different subclones. Based on the results, kataegis most likely occurs in different subclones across the tumor's evolution; however, this conclusion need to be confirmed. The remaining questions concern (2) whether the genomic alterations derived from kataegis or chromothripsis are enriched with cancer drivers; (3) whether kataegis or chromothripsis results from specifically defective internal cellular machinery or certain combinatory patterns of existing mutations or from other epigenetic alterations; (4) whether katae-

Table 1

A list of major completed tumor genome sequencing projects.

Tumor type	Number of tumors sequenced	Year	Ref.
Lung cancer	31	2012	[41]
Lung cancer	53	2012	[42]
Lung cancer	183	2012	[43]
Lung cancer	178	2012	[44]
Breast cancer	825	2012	[45]
Breast cancer	125	2012	[46]
Breast cancer	77	2012	[47]
Breast cancer	21	2012	[7]
Breast cancer	100	2012	[48]
Breast cancer	65	2012	[49]
Breast cancer	11	2006	[32]
Medulloblastoma	37	2012	[4]
Medulloblastoma	92	2012	[50]
Medulloblastoma	125	2012	[5]
Acute myeloid leukemia	24	2012	[51]
Multiple myeloma	38	2011	[52]
Chronic lymphocytic leukemia	105	2012	[53]
Melanoma	121	2012	[54]
Melanoma	25	2012	[55]
Prostate cancer	112	2012	[56]
Prostate cancer	50	2012	[57]
Prostate cancer	23	2011	[58]
Colon and rectal cancer	224	2012	[59]
Colon cancer	11	2006	[32]
Liver cancer	27	2012	[60]
Hepatocellular carcinoma	24	2012	[61]
Diffuse large B-cell lymphoma	55	2012	[62]
Pediatric glioblastoma	48	2012	[63]
Gastric cancer	22	2011	[64]
Ovarian carcinoma	316	2011	[65]
Head and neck carcinoma	74	2011	[66]
Head and neck carcinoma	32	2011	[67]

Note: Projects that have sequenced more than 20 cancer genomes at an exome or an entire genome level have been selected. The majority of the work was published in 2012.

gis-derived cancer genomes are more aggressive than the tumors derived from other genomic alteration mechanisms; and (5) whether the DNA copy number change represents an important genetic alteration in cancer cells. A series of chromothripsis might convert a cell into a cancer cell quickly. However, these cells may not mutate but rather undergo many copy number changes. Further investigations of these novel genomic alteration mechanisms may deepen the understanding of factors influencing these mutations/copy number variations and cancer development. Nevertheless, early mutations, copy number changes or alterations of chromosomal structures caused by other factors, such as smoking, sunlight, or stress, may be conducive to kataegis or chromothripsis.

3. Genomic alterations and cancer genome sequencing

3.1. Deciphering cancer genomic alterations in an evolutionary context

Tumors are widely known to be heterogeneous. Therefore, a tumor is often expected to comprise a set of clones of cancer cells. Understanding the mechanisms that cause cancer requires comprehensive insight into the processes underlying cancer genetics. To obtain this information, the Sanger group sequenced 21 breast cancer genomes, cataloged their mutations and attempted to trace the evolution of cancer mutation profiles between different subclones in the context of timing [7]. Using appropriate bioinformatic approaches, they were able to uncover the genetic mutation marks and the inscribed record of the latter's emergence over time, thus enabling the divergence of a cell in forming the various clones to be traced.

Cancer mutational processes evolve over a tumor's lifespan. Each time a gene is mutated in a cell, the mutation event encrypts a distinctive stamp inscribed in its genome. Such a distinctive stamp can be considered a mutation signature or a distinctive mark on the genome. From a historical perspective, an archaeological history of mutations results from the accumulations of marks upon one another. Therefore, collecting, cataloging and sorting the orders of these marks allow the cancer's life history to be deciphered.

Cells in a single subclone possess an identical genetic mutation profile. The cells in late-occurring subclones add new mutations to the genetic mutation profiles of the early subclones. Only a mutation causes a clone to dramatically expand, after which the clone constitutes the tumor's dominant population and the tumor grows large enough to be diagnosed. During this process, the history of the genetic mutations and the emergence and the expansion of the new clone is imprinted upon the cancer genome. Therefore, these genetic mutation profiles can be used to build an evolutionary tree of subclones. This analysis allows the tracing of the occurrence of a new mutation profile, the biological processes involved in each mutation profile or subclone, and the proportion of each subclone present in the tumors. The study showed that as the mutations, or genetic marks, accumulate in a cell, the corresponding changes caused the cancer to diverge into different clones [7]. The emergence of different subclones occurred in all of the examined cancer genomes. This analysis suggests that when a new subclone has grown sufficiently to constitute the dominant population of cancer cells (i.e., more than 50% of tumor cells), the tumor becomes clinically detectable. The genetic mutation profiles of subclones and dominant subclones differ markedly, implying that the dominant subclone may require a significant amount of time to develop.

This study has several implications. First, it confirms that a tumor contains many subclones; thus, the heterogeneity can be sampled by sequencing, as shown in Gerlinger et al. [7]. Second, the study suggests that a cancer patient may benefit from the use of combinatory drugs for different subclones. One perspective is that

treatment may proceed by targeting the early clones because their mutation profiles are identical to those of the late subclones [8]. However, this approach might not work well. The late subclones, particularly the dominant subclones, acquire new traits for cell proliferation and survival. These new traits could grant the late subclones "new" survival capabilities, thus implying that the 'drug targets' for the early subclones are unsuitable for the late subclones. However, subclones [9] can successfully be dissected from tumors, and the RNA-seq or sequencing analysis of these subclones can be performed [10]. Profiling these subclones helps to identify drug targets for each subclone. Third, each cancer appears to be generated by a distinct combination of mutational processes. Fourth, the study confirms the prediction that transforming a normal cell to a cancer cell is a lengthy process that has been shown to last 10 years [11]. This fact supports the search for mutation signatures or molecular biomarkers for early diagnosis and the possibility of personalized early preventions to potential cancer patients.

3.2. Which subclone contains the cancer stem cell?

Although many implications have been found from deciphering cancer genomes, a key question remains: can we identify or infer which subclone contains the cancer stem cell?

The cancer stem cell hypothesis suggests that tumors are organized into aberrant cell hierarchies. In such a hierarchy, a cancer stem cell capable of replicating indefinitely is the parent cell that generates differentiated daughter cells with a limited capacity to proliferate. If proven for a tumor, this hypothesis suggests that cancer is curable if drugs are available to kill cancer stem cells. Recently, three studies have provided the first evidence that cancer stem cells exist [12]. These studies have demonstrated that cancer stem cells reproduce cancer cells when the tumor is extinguished by anticancer drugs, a result that applies to benign and malignant tumors.

These results suggest that cancer stem cells are inherently capable of propagating genetic or epigenetic alterations throughout a tumor and driving cancer evolution. The cancer stem cells constitute the stem of tumor cell production. During a cancer's evolution, a certain genetic mutation profile may define cancer stem cell subclones. The parental subclones of the cancer stem cell subclone are non-stem cells, thus indicating that genetic/epigenetic changes could convert a non-stem cell into a cancer stem cell. Such a stem cell produces many daughter cells that may further evolve into diversified cancer cells. If this process occurs, then the dominant subclones of the tumors would not be cancer stem cells but would emerge from cancer stem cells.

Thus, many questions remain. First, should every tumor contain a subset of cancer stem cells? All tumors may not contain cancer stem cells, for example, if a chromothripsis occurs to generate enough fitting combinations of mutations to generate a new subclone capable of becoming dominant in a tumor. However, generally, sustaining the cancer's evolution requires generating enough cancer cells using cancer stem cells. In this case, a cancer stem cell must evolve and be defined by a set of mutations. Second, if a cancer stem cell must exist in a tumor, what are the characteristics of its genetic mutation profiles? The evolution of cancer genomes implies that a cancer stem cell subclone should be an early subclone in the history of the tumor's development. We assume that the cancer stem cell subclone generates new cells but at a limited and slow rate because of the microenvironment's unfavorable conditions. These daughter cells are further mutated and generate new subclones.

This hypothesis is supported by the discoveries of Driessens et al. [12–14] and Chen et al. [13] that show that the cellular organization of early (precancerous) skin and intestinal tumors are composed of both stem cells and non-stem cells. Furthermore,

Driessens et al. observed that progression to cancer in benign skin tumors is associated with an increased cancer stem cell population. They found that every cancer stem cell within a tumor is equally probable of clonally expanding or dying, possibly even in the absence of new mutations, thus suggesting that a cancer stem cell subclone most likely occurs in an early stage of cancer development (Fig. 1).

The most important finding from the above discussion is that subclonal diversity in cancer genomes may be dissected. To treat cancer more efficiently, it requires the dissection of not only the dominant subclones but also the minor subclones, which are among the top subclones in the hierarchy and most likely cancer stem cells. The key challenges are determining the functions of the mutations in these subclones and linking these mutations to drug targets for each subclone and, subsequently, identifying mutational signatures as diagnostic tools that predict the early development of cancer.

4. The functional understanding of genomic alterations in cancer genomes at a systems level

As discussed above, in recent years, sequencing efforts have generated many genomic alterations of cancer. Although some important insights, such as new genomic alteration mechanisms, cancer genome evolution, and clonal dissection, have been gleaned from the corresponding data, a lack of frameworks and tools nevertheless persists with respect to how genomic alterations drive cancer evolution, clonal expansion, and cancer development and metastasis. Understanding these processes requires investigating the function of these genomic alterations. However, currently, the greatest challenge is understanding the functions of driver mutations in cancer genomes. On average, several hundred driver mutations have been cataloged from each tumor, but interpreting these mutations is inefficient, thus limiting the understanding of why and how these mutation genes work. Most current cancer genome sequencing work has mapped the mutations onto biological pathways, including signaling pathways.

A biological pathway contains a set of linear relationships between genes or proteins and is an important model to describe biological relationships. From an abstract perspective, a pathway is a model that represents humans' efforts to explore, describe, and organize biological relationships in cells [15]. Such efforts allow us to further understand biological systems and to predict cell behaviors. However, less than 5–10% of mutations can be explained by these 'core pathways' in which a set of genes is relatively highly mutated [16,17]. Several problems are associated with this approach. First, pathway gene members are fuzzy and poorly defined. For example, if different databases such as KEGG or BioCarta are queried for 'EGFR signaling pathway,' each database will provide a different gene list. Second, only a small set of muta-

tion genes (less than 5–10%) can be mapped onto the 'core pathways,' thus, most mutation genes are unmapped. Third, for each study, only a few 'core pathways' such as RAS-PI3K were found repeatedly. These pathways are well known in cancer studies, and drugs have been developed for these pathways or key genes. In this respect, the pathway approach does not provide new insights but captures the existing knowledge in the field.

To comprehensively understand the drivers of cancer genomes, computational methods should be developed because of the large number of driver mutation genes in each tumor. The development of effective computational methods first requires understanding what subjects biology addresses. Over many years, biologists have explored several types of relationships among genes, proteins, RNAs and other molecules. The molecular mechanisms of biological processes or diseases have also been studied. A molecular mechanism often discloses an event of A regulating B, B inducing or inhibiting C and generating a phenotype or disease. Each mechanism comprises a set of relationships between molecules such as gene regulation, protein interaction, activation, genetic interaction and inhibitory action. In this context, biology is a science that describes relationships [15].

Traditionally, biologists use pathways to describe the relationships between a limited number of genes. As shown above, numerous driver mutations have been found in each tumor, but the pathway approach explains only 5–10% of the genetic mutations in the cancer genome. It is expected that biologists must consider thousands of biological relationships with respect to these genes in one tumor, for which task the traditional methods of describing biological relationships are inadequate. Furthermore, as the cross-talk between pathways has been documented extensively in recent years, scientists have come to recognize that a network model may provide a solution. A network model describes the interactions and relationships between genes or proteins in a non-linear manner. The pathway and network approaches are conceptual models that reflect an avenue of exploring biology by humans. Both models project living cell molecules onto conceptual frameworks, but the molecular network model more closely approximates the processes that occur in cells.

4.1. Basic concepts of molecular networks

A molecular network can be represented using nodes and links. Nodes represent genes or proteins, whereas links represent the relationships between genes and proteins. Cancer omic data can be integrated into a molecular network to pose questions and obtain corresponding insights. However, different types of networks have distinct properties and thus can be used to address different questions.

There are four basic molecular networks: signaling, gene regulatory, metabolic and protein interaction networks. A network hub consists of the nodes that have the most links (top 5–10% of

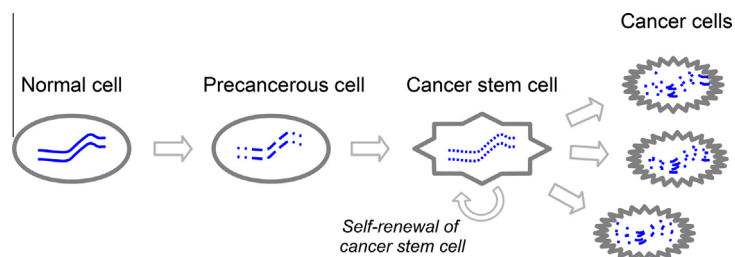


Fig. 1. A hypothesis of a stem cell-based cancer genome evolution. The production of a cancer stem cell requires a long time (5–10 years) and a series of changes with respect to somatic mutations, copy number variations and epigenetic factors from a normal cell. The cancer stem cell generates cancer cells. Some mutations or copy number changes could lead one cancer cell derived from the cancer stem cell to acquire significant growth advantages. This cancer cell will propagate to dominate the tumor and eventually make the tumor clinically detectable.

the links) within the network. However, a hub assumes a different biological meaning in different networks: hubs in gene regulatory networks represent global transcription factors, which assume a major role in responding to stimuli and coordinating the regulated genes. Consistent with this mechanism, the transcripts of the hub transcription factors often rapidly decay [18]. The rapidly decaying transcripts of global hub transcription factors might encode “switch” functions, which are used under different conditions and stimuli. Hubs in cellular signaling and metabolic networks help to organize information flows within the networks. Hubs in signaling networks assume a major role in integrating different signals and pathways [19]. However, hubs in protein interactions indicate that hub proteins are involved in considerably more biological processes/protein complexes than other proteins [20].

Most targets of transcription factors in regulatory networks are “workers.” They directly execute the tasks of biological processes. Most nodes in gene regulatory network have no regulatory role, whereas most nodes in cellular signaling networks have regulatory roles. In signaling networks, approximately all nodes are “regulators.” Therefore, logical regulatory relationships are extensively encoded in signaling networks. In another view, nodes in signaling networks are information carriers. Similarly, nodes in metabolic networks are carriers of the metabolic flux. Nodes in protein interaction networks are members of protein complexes, or ‘co-workers,’ that interact and collaborate to complete specific tasks [15].

Generally, signaling networks contain several logical codes of regulation, whereas protein interaction networks do not encode regulatory logics. Gene regulatory networks encode both regulatory logic and gene “workers.” Gene regulatory networks help to identify key cancer regulators, the co-expression of genes, and the sets of “workers” involved in cancer processes, whereas signaling networks help to identify cancer causal genes and the regulatory logic in cancer processes [15]. Protein interaction networks provide a source for identifying co-workers in cancer progression or metastasis [15]. In recent years, notable discoveries have also addressed cancer metabolism. For example, hypoxia and isocitrate dehydrogenase 1 (*IDH1*) function in the citric acid cycle pathway, an energy-generating pathway. The *IDH1* mutations cause the accumulation of 2-hydroxyglutarate, a metabolite in the metabolic pathway, and thus cause cancer [21–23]. These studies generate interest in studying cancer in metabolic networks. Metabolic networks could help identify key enzymes or metabolites that could induce metabolic chaos and thus cause cancer.

When incorporating cancer omic data into these networks, the expression levels of major regulators (i.e., kinases) in signaling networks do not necessarily change dramatically during cancer progression and metastasis because signaling cascades are modulated through protein modifications, not gene regulation. Indeed, examining phosphoproteomic data in signaling networks is informative [24].

4.2. An integrative network analysis of cancer omic data

Prior to the advent of cancer genome sequencing, researchers pursued network analyses of cancer-related data. Additional details regarding this research are available in reviews [25,26] and in the book *Cancer Systems Biology* [15]. For example, an integrative analysis of tumor gene expression profiles and protein interaction networks permits the identification of network modules as cancer biomarkers. A network module contains a set of closely interacted genes/proteins that collaborate to perform certain biological process. In a network module, the interactions between the module members are more frequent than the interactions between the module members with the protein/nodes outside the module. The resulting network module markers are more reproducible than

traditionally obtained biomarkers that are identified without using protein interaction networks [27,28].

Another study found that the protein network modules of gene co-expression are more reliable predictors than the biomarkers that are identified without considering network modules [29]. Exhaustively describing human molecular networks is a long-term task. Therefore, although the network module approach has yielded promising results, the accuracy of its predictions is nevertheless low (less than 80%). However, excellent results (i.e., over 90% prediction accuracy when tested in multiple independent patient cohorts) have been obtained with biomarkers through cancer-hallmark-based functional modules that extend beyond the network module [30].

A functional module biomarker is linked to a network module of cancer driver-mutating genes on the protein interaction network. Notably, as long as the genes, which perform the same function of the module and act as the interacting partners of proteins in the cancer-driver network module, are modulated, they show predictive power (Fig. 2) [30]. Most importantly, the members of the cancer-driver network module are fixed, but the modules’ modulated interacting partners change with patient cohorts [30]. These results explain why many potential biomarker sets exhibit limited overlap, a well-known phenomenon in the cancer biomarker field.

The first study that attempted to interpret the drivers behind cancer-mutating genes in networks was published in 2007. In this work, Cui et al. [31] first examined the traditionally discovered cancer-mutating genes, which were considered to be cancer drivers and which were mined using the COSMIC database, using a manually curated human signaling network. They identified 12 cancer signaling network modules in which cancer-mutating genes were densely connected. Such modules cannot be discovered using a pathway approach or by examining mutating genes individually. Cui et al. applied these modules to the drivers behind the mutating genes derived from the first genome-wide sequencing of cancer [32] and found that these mutations became meaningful if assigned to 2 or 3 signaling network modules [31].

This result resolved the controversial arguments concerning whether tumor genome sequencing should continue [33] because the first study of tumor genome sequencing showed that cancer mutations are highly complex: driver-mutating genes are rare, even in the tumors of sample origin, such as breast cancer [34]. In addition, these genes are not organized into pathways. Subsequently, some scientists argued that cancer genome sequencing is time- and resource-consuming and provides unclear results [33]. Cui et al.’s network analysis provides evidence that cancer driver-mutating genes are organized into network modules to assume roles. Lincoln Stein and colleagues [35] used an extended signaling network to examine cancer driver-mutating genes derived from the genome sequencing of two glioblastoma multiforme data sets [17,36] and confirmed that cancer driver-mutating genes are organized into network modules.

Given that current networks are incomplete, Ciriello et al. [37] used correlation analysis and statistical approaches to develop a method for identifying cancer gene modules, specifically, Mutual Exclusivity Modules in cancer (MEMo). MEMo considers the frequency of cancer gene mutation across a set of tumor samples. Cancer mutation genes are most likely to participate in the same biological process, and cancer-mutating genes that are involved in the same biological process are mutually exclusive in a module. This approach may help detect cancer gene modules that cannot be found with existing networks. However, a limitation of the method is that the sample number significantly affects the results.

Akavia et al. [38] used a Bayesian network approach to develop a network-module-based method, CONEXIC (computational algorithm, copy number and expression in cancer) to examine DNA

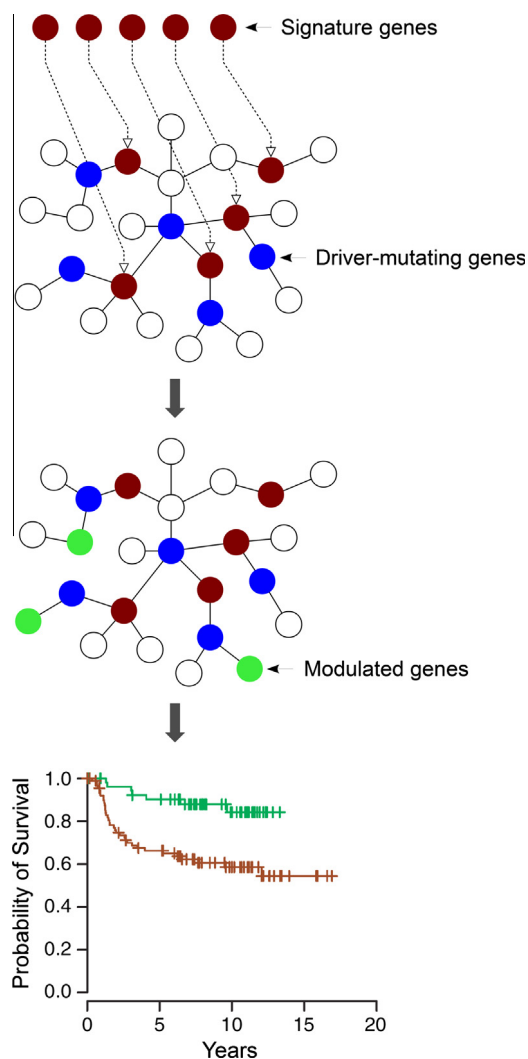


Fig. 2. Stable cancer driver-mutating gene modules have predictive power in cancer prognosis. The content of this figure was developed from a cancer prognostic gene signature derived from Li et al. [30]. For example, this gene signature represents apoptosis, one of the cancer hallmarks. A driver-module is formed by collecting the genes (blue) that are the interacting neighbors of the signature genes (brown). By obtaining microarray gene expression profiles from a breast cancer cohort, we obtained modulated genes between recurred and non-recurred tumors. We sought a predictive gene signature by collecting the genes (green) that are not only the interacting neighbors of the driver-module genes (blue), but also the modulated genes sharing the same function with the original gene signature (i.e., apoptosis). We found that the predictive gene signature (green genes) has predictive power for prognosis. It is notable that another set of the predictive gene signature (green genes) can be obtained when changing patient cohorts. However, the new genes in green maintain their predictive power. These results explain the phenomenon that the gene sets of different cancer biomarkers for an identical cancer type rarely overlap. More detailed explanations are in Li et al. [30]. Protein interaction networks contain circles (nodes) and solid lines that represent genes and their interactions, respectively. Brown, blue and green nodes represent signature genes, breast cancer driver-mutating genes, and breast cancer modulated genes, respectively, between recurred and non-recurred tumors. Dotted lines represent the signature genes mapped to the genes in the network. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

copy number alterations in tumor genomes. This approach integrates copy number alterations and gene expression profiles to identify potential cancer driver genes that have recurrent copy number changes in tumor samples. A limitation of this study is that the Bayesian network learning approach and other network learning approaches exhibit weak performance for eukaryotic genomes, such as those in yeast and humans [39].

5. Future directions

An integrative network analysis of cancer genome sequencing data is in its early phases. Several challenges and questions need to be addressed. First, in terms of their functional importance, the cancer driver-mutating genes need to be prioritized into categories such as cancer cell survival or metastasis. This task is urgent because each tumor contains several hundred cancer driver-mutating genes, but it is unclear which genes should be prioritized in research. New network methods can be developed to answer this question. Second, the network of cancer stem cells requires further study to understand which combinations of mutating genes or epigenetic/copy number changes maintain the characteristics of cancer stem cells. Third, the network dynamics of cancer evolution should be mapped, and the key events that lead to cancer initiation need to be understood. Fourth, the networks of cancer hallmarks should be modeled. This approach has been proposed, and its necessity has been detailed in the book *Cancer Systems Biology* [15]. Li et al. obtained excellent results in identifying cancer biomarkers by modeling functional modules of cancer hallmarks [30]. Fu et al. also showed that apoptotic network motif modeling based on one of the cancer hallmarks inferred that the 26S proteasome may help predict cancer prognosis [40]. Fifth, the dynamics of the cancer gene network motifs need to be studied, and the regulatory principles and controls of the oncogenic signaling information flows should be understood. In-depth discussions are available in a recent review [26]. Finally, new network algorithms need to be developed to translate some cancer mutations and other genetic/epigenetic changes into drug targets. Inarguably, many other problems and activities may be solved through an integrative network approach. With additional cancer high-throughput studies in the near future [68], we believe that the integrative network approach will become increasingly useful to cancer research.

Acknowledgements

This work is supported by Canadian Institute of Health Research and NRC's Genome and Health Initiative.

References

- [1] Y. Joly, E.S. Dove, B.M. Knoppers, M. Bobrow, D. Chalmers, Data sharing in the post-genomic world: the experience of the international cancer genome consortium (ICGC) data access compliance office (DACO), *PLoS. Comput. Biol.* 8 (2012) e1002549.
- [2] J.R. Downing, R.K. Wilson, J. Zhang, E.R. Mardis, C.H. Pui, L. Ding, T.J. Ley, W.E. Evans, The pediatric cancer genome project, *Nat. Genet.* 44 (2012) 619–622.
- [3] P.J. Stephens, C.D. Greenman, B. Fu, F. Yang, G.R. Bignell, L.J. Mudie, E.D. Pleasance, K.W. Lau, D. Beare, L.A. Stebbings, S. McLaren, M.L. Lin, D.J. McBride, I. Varela, S. Nik-Zainal, C. Leroy, M. Jia, A. Menzies, A.P. Butler, J.W. Teague, M.A. Quail, J. Burton, H. Swerdlow, N.P. Carter, L.A. Morsberger, C. Iacobuzio-Donahue, G.A. Follows, A.R. Green, A.M. Flanagan, M.R. Stratton, P.A. Futreal, P.J. Campbell, Massive genomic rearrangement acquired in a single catastrophic event during cancer development, *Cell* 144 (2011) 27–40.
- [4] G. Robinson, M. Parker, T.A. Kranenburg, C. Lu, X. Chen, L. Ding, T.N. Phoenix, E. Hedlund, L. Wei, X. Zhu, N. Chalhoub, S.J. Baker, R. Huether, R. Kriwacki, N. Curley, R. Thiruvakatam, J. Wang, G. Wu, M. Rusch, X. Hong, J. Becksfors, P. Gupta, J. Ma, J. Easton, B. Vadodaria, A. Onar-Thomas, T. Lin, S. Li, S. Pounds, S. Paugh, D. Zhao, D. Kawachi, M.F. Roussel, D. Finkelstein, D.W. Ellison, C.C. Lau, E. Bouffet, T. Hassall, S. Gururangan, R. Cohn, R.S. Fulton, L.L. Fulton, D.J. Dooling, K. Ochoa, A. Gajjar, E.R. Mardis, R.K. Wilson, J.R. Downing, J. Zhang, R.J. Gilbertson, Novel mutations target distinct subgroups of medulloblastoma, *Nature* 488 (2012) 43–48.
- [5] D.T. Jones, N. Jager, M. Kool, T. Zichner, B. Hutter, M. Sultan, Y.J. Cho, T.J. Pugh, V. Hovestadt, A.M. Stutz, T. Rausch, H.J. Warnatz, M. Ryzhova, S. Bender, D. Sturm, S. Pleier, H. Cin, E. Pfaff, L. Sieber, A. Wittmann, M. Remke, H. Witt, S. Hutter, T. Tzaridis, J. Weischenfeldt, B. Raeder, M. Avci, V. Amstislavskiy, M. Zapatka, U.D. Weber, Q. Wang, B. Lasitschka, C.C. Bartholomae, M. Schmidt, K.C. von, V. Ast, C. Lawrenz, J. Eils, R. Kabbe, V. Benes, S.P. van, J. Koster, R. Volckmann, D. Shih, M.J. Betts, R.B. Russell, S. Coco, G.P. Tonini, U. Schuller, V. Hans, N. Graf, Y.J. Kim, C. Monoranu, W. Roggendorf, A. Unterberg, C. Herold-Mende, T. Milde, A.E. Kulozik, D.A. von, O. Witt, E. Maass, J. Rossler, M. Ebinger, M.U. Schuhmann, M.C. Fruhwald, M. Hasselblatt, N. Jabado, S. Rutkowski, A.O.

- von Bueren, D. Williamson, S.C. Clifford, M.G. McCabe, V.P. Collins, S. Wolf, S. Wiemann, H. Lehrach, B. Brors, W. Scheurlen, J. Felsberg, G. Reifenberger, P.A. Northcott, M.D. Taylor, M. Meyerson, S.L. Pomeroy, M.L. Yaspo, J.O. Korbel, A. Korshunov, R. Eils, S.M. Pfister, and P. Lichter, Dissecting the genomic complexity underlying medulloblastoma, *Nature* 488 (2012) 100–105.
- [6] S. Nik-Zainal, L.B. Alexandrov, D.C. Wedge, L.P. Van, C.D. Greenman, K. Raine, D. Jones, J. Hinton, J. Marshall, L.A. Stebbings, A. Menzies, S. Martin, K. Leung, L. Chen, C. Leroy, M. Ramakrishna, R. Rance, K.W. Lau, L.J. Mudie, I. Varela, D.J. McBride, G.R. Bignell, S.L. Cooke, A. Shlien, J. Gamble, I. Whitmore, M. Maddison, P.S. Tarpey, H.R. Davies, E. Papaemmanuil, P.J. Stephens, S. McLaren, A.P. Butler, J.W. Teague, G. Jonsson, J.E. Garber, D. Silver, P. Miron, A. Fatima, S. Boyault, A. Langerod, A. Tutt, J.W. Martens, S.A. Aparicio, A. Borg, A.V. Salomon, G. Thomas, A.L. Borresen-Dale, A.L. Richardson, M.S. Neuberg, P.A. Futreal, P.J. Campbell, M.R. Stratton, Mutational processes molding the genomes of 21 breast cancers, *Cell* 149 (2012) 979–993.
- [7] S. Nik-Zainal, L.P. Van, D.C. Wedge, L.B. Alexandrov, C.D. Greenman, K.W. Lau, K. Raine, D. Jones, J. Marshall, M. Ramakrishna, A. Shlien, S.L. Cooke, J. Hinton, A. Menzies, L.A. Stebbings, C. Leroy, M. Jia, R. Rance, L.J. Mudie, S.J. Gamble, P.J. Stephens, S. McLaren, P.S. Tarpey, E. Papaemmanuil, H.R. Davies, I. Varela, D.J. McBride, G.R. Bignell, K. Leung, A.P. Butler, J.W. Teague, S. Martin, G. Jonsson, O. Mariani, S. Boyault, P. Miron, A. Fatima, A. Langerod, S.A. Aparicio, A. Tutt, A.M. Sieuwerts, A. Borg, G. Thomas, A.V. Salomon, A.L. Richardson, A.L. Borresen-Dale, P.A. Futreal, M.R. Stratton, P.J. Campbell, The life history of 21 breast cancers, *Cell* 149 (2012) 994–1007.
- [8] T.A. Yap, M. Gerlinger, P.A. Futreal, L. Pusztai, C. Swanton, Intratumor heterogeneity: seeing the wood for the trees, *Sci. Transl. Med.* 4 (2012) 127ps10.
- [9] D. Ramskold, S. Luo, Y.C. Wang, R. Li, Q. Deng, O.R. Faridani, G.A. Daniels, I. Khrebtkova, J.F. Loring, L.C. Laurent, G.P. Schroth, R. Sandberg, Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells, *Nat. Biotechnol.* (2012).
- [10] S.L. Carter, K. Cibulskis, E. Helman, A. McKenna, H. Shen, T. Zack, P.W. Laird, R.C. Onofrio, W. Winckler, B.A. Weir, R. Beroukhim, D. Pellman, D.A. Levine, E.S. Lander, M. Meyerson, G. Getz, Absolute quantification of somatic DNA alterations in human cancer, *Nat. Biotechnol.* 30 (2012) 413–421.
- [11] S. Yachida, S. Jones, I. Bozic, T. Antal, R. Leary, B. Fu, M. Kamiyama, R.H. Hruban, J.R. Eshleman, M.A. Nowak, V.E. Velculescu, K.W. Kinzler, B. Vogelstein, C.A. Iacobuzio-Donahue, Distant metastasis occurs late during the genetic evolution of pancreatic cancer, *Nature* 467 (2010) 1114–1117.
- [12] G. Driessens, B. Beck, A. Caauwe, B.D. Simons, C. Blanpain, Defining the mode of tumour growth by clonal analysis, *Nature* 488 (2012) 527–530.
- [13] J. Chen, Y. Li, T.S. Yu, R.M. McKay, D.K. Burns, S.G. Kernie, L.F. Parada, A restricted cell population propagates glioblastoma growth after chemotherapy, *Nature* 488 (2012) 522–526.
- [14] A.G. Schepers, H.J. Snippert, D.E. Stange, M. van den Born, J.H. van Es, M. van de Wetering, H. Clevers, Lineage tracing reveals Lgr5+ stem cell activity in mouse intestinal adenomas, *Science* 337 (2012) 730–735.
- [15] E. Wang, *Cancer Systems Biology*, CRC Press, Boca Raton, FL, 2010.
- [16] S. Jones, X. Zhang, D.W. Parsons, J.C. Lin, R.J. Leary, P. Angenendt, P. Mankoo, H. Carter, H. Kamiyama, A. Jimeno, S.M. Hong, B. Fu, M.T. Lin, E.S. Calhoun, M. Kamiyama, K. Walter, T. Nikolskaya, Y. Nikolsky, J. Hartigan, D.R. Smith, M. Hidalgo, S.D. Leach, A.P. Klein, E.M. Jaffe, M. Goggins, A. Maitra, C. Iacobuzio-Donahue, J.R. Eshleman, S.E. Kern, R.H. Hruban, R. Karchin, N. Papadopoulos, G. Parmigiani, B. Vogelstein, V.E. Velculescu, K.W. Kinzler, Core signaling pathways in human pancreatic cancers revealed by global genomic analyses, *Science* 321 (2008) 1801–1806.
- [17] Cancer Genome Atlas Research Network, Comprehensive genomic characterization defines human glioblastoma genes and core pathways, *Nature* 455 (2008) 1061–1068.
- [18] E. Wang, E. Purisima, Network motifs are enriched with transcription factors whose transcripts have short half-lives, *Trends Genet.* 21 (2005) 492–495.
- [19] A. Awan, H. Bari, F. Yan, S. Moksong, S. Yang, S. Chowdhury, Q. Cui, Z. Yu, E.O. Purisima, E. Wang, Regulatory network motifs and hotspots of cancer genes in a mammalian cellular signalling network, *IET. Syst. Biol.* 1 (2007) 292–297.
- [20] Q. Cui, E.O. Purisima, E. Wang, Protein evolution on a human signaling network, *BMC. Syst. Biol.* 3 (2009) 21.
- [21] P. Koivunen, S. Lee, C.G. Duncan, G. Lopez, G. Lu, S. Ramkissoon, J.A. Losman, P. Joensuu, U. Bergmann, S. Gross, J. Travins, S. Weiss, R. Looper, K.L. Ligon, R.G. Verhaak, H. Yan, W.G. Kaelin Jr., Transformation by the (R)-enantiomer of 2-hydroxyglutarate linked to EGLN activation, *Nature* 483 (2012) 484–488.
- [22] C. Lu, P.S. Ward, G.S. Kapoor, D. Rohle, S. Turcan, O. Abdel-Wahab, C.R. Edwards, R. Khanin, M.E. Figueroa, A. Melnick, K.E. Wellen, D.M. O'Rourke, S.L. Berger, T.A. Chan, R.L. Levine, I.K. Mellingshoff, C.B. Thompson, IDH mutation impairs histone demethylation and results in a block to cell differentiation, *Nature* 483 (2012) 474–478.
- [23] S. Turcan, D. Rohle, A. Goenka, L.A. Walsh, F. Fang, E. Yilmaz, C. Campos, A.W. Fabius, C. Lu, P.S. Ward, C.B. Thompson, A. Kaufman, O. Guryanova, R. Levine, A. Heguy, A. Viale, L.G. Morris, J.T. Huse, I.K. Mellingshoff, T.A. Chan, IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype, *Nature* 483 (2012) 479–483.
- [24] L. Li, C. Tibiche, C. Fu, T. Kaneko, M.F. Moran, M.R. Schiller, S.S. Li, E. Wang, The human phosphotyrosine signaling network: evolution and hotspots of hijacking in cancer, *Genome Res.* 22 (2012) 1222–1230.
- [25] E. Wang, A. Lenferink, M. O'Connor-McCourt, Cancer systems biology: exploring cancer-associated genes on cellular networks, *Cell Mol. Life Sci.* 64 (2007) 1752–1762.
- [26] M. Cloutier, E. Wang, Dynamic modeling and analysis of cancer cellular network motifs, *Integr. Biol. (Camb.)* 3 (2011) 724–732.
- [27] H.Y. Chuang, E. Lee, Y.T. Liu, D. Lee, T. Ideker, Network-based classification of breast cancer metastasis, *Mol. Syst. Biol.* 3 (2007) 140.
- [28] M. Paliouras, N. Zaman, R. Lumbroso, L. Kapogeorgakis, L.K. Beitel, E. Wang, M. Trifiro, Dynamic rewiring of the androgen receptor protein interaction network correlates with prostate cancer clinical outcomes, *Integr. Biol. (Camb.)* 3 (2011) 1020–1032.
- [29] I.W. Taylor, R. Linding, D. Warde-Farley, Y. Liu, C. Pesquita, D. Faria, S. Bull, T. Pawson, Q. Morris, J.L. Wrana, Dynamic modularity in protein interaction networks predicts breast cancer outcome, *Nat. Biotechnol.* 27 (2009) 199–204.
- [30] J. Li, A.E. Lenferink, Y. Deng, C. Collins, Q. Cui, E.O. Purisima, M.D. O'Connor-McCourt, E. Wang, Identification of high-quality cancer prognostic markers and metastasis network modules, *Nat. Commun.* 1 (2010) 34.
- [31] Q. Cui, Y. Ma, M. Jaramillo, H. Bari, A. Awan, S. Yang, S. Zhang, L. Liu, M. Lu, M. O'Connor-McCourt, E.O. Purisima, E. Wang, A map of human cancer signaling, *Mol. Syst. Biol.* 3 (2007) 152.
- [32] T. Sjoblom, S. Jones, L.D. Wood, D.W. Parsons, J. Lin, T.D. Barber, D. Mandelker, R.J. Leary, J. Ptak, N. Silliman, S. Szabo, P. Buckhaults, C. Farrell, P. Meeh, S.D. Markowitz, J. Willis, D. Dawson, J.K. Willson, A.F. Gazdar, J. Hartigan, L. Wu, C. Liu, G. Parmigiani, B.H. Park, K.E. Bachman, N. Papadopoulos, B. Vogelstein, K.W. Kinzler, V.E. Velculescu, The consensus coding sequences of human breast and colorectal cancers, *Science* 314 (2006) 268–274.
- [33] W.J. Chng, Limits to the Human Cancer Genome Project?, *Science* 315 (2007) 762–765.
- [34] J. Kaiser, Cancer. First pass at cancer genome reveals complex landscape, *Science* 311 (3) (2006) 1370.
- [35] G. Wu, X. Feng, L. Stein, A human functional protein interaction network and its application to cancer data analysis, *Genome Biol.* 11 (2010) R53.
- [36] D.W. Parsons, S. Jones, X. Zhang, J.C. Lin, R.J. Leary, P. Angenendt, P. Mankoo, H. Carter, I.M. Siu, G.L. Gallia, A. Olivi, R. McLendon, B.A. Rasheed, S. Keir, T. Nikolskaya, Y. Nikolsky, D.A. Busam, H. Tekleab, L.A. Diaz Jr., J. Hartigan, D.R. Smith, R.L. Strausberg, S.K. Marie, S.M. Shinjo, H. Yan, G.J. Riggins, D.D. Bigner, R. Karchin, N. Papadopoulos, G. Parmigiani, B. Vogelstein, V.E. Velculescu, K.W. Kinzler, An integrated genomic analysis of human glioblastoma multiforme, *Science* 321 (2008) 1807–1812.
- [37] G. Ciriello, E. Cerami, C. Sander, N. Schultz, Mutual exclusivity analysis identifies oncogenic network modules, *Genome Res.* 22 (2012) 398–406.
- [38] U.D. Akavia, O. Litvin, J. Kim, F. Sanchez-Garcia, D. Kotliar, H.C. Causton, P. Pochanard, E. Mozes, L.A. Garraway, D. Pe'er, An integrated approach to uncover drivers of cancer, *Cell* 143 (2010) 1005–1017.
- [39] D. Marbach, J.C. Costello, R. Kuffner, N.M. Vega, R.J. Prill, D.M. Camacho, K.R. Allison, M. Kellis, J.J. Collins, G. Stolovitzky, Wisdom of crowds for robust gene network inference, *Nat. Methods* 9 (2012) 796–804.
- [40] C. Fu, J. Li, E. Wang, Signaling network analysis of ubiquitin-mediated proteins suggests correlations between the 26S proteasome and tumor progression, *Mol. Biosyst.* 5 (2009) 1809–1816.
- [41] M. Peifer, L. Fernandez-Cuesta, M.L. Sos, J. George, D. Seidel, L.H. Kasper, D. Plenker, F. Leenders, R. Sun, T. Zander, R. Menon, M. Koker, I. Dahmen, C. Muller, C. Di, V. H.U. Schildhaus, J. Altmuller, I. Baessmann, C. Becker, W.B. de, J. Vandesompele, D. Bohm, S. Ansen, F. Gabler, I. Wilkening, S. Heynck, J.M. Heuckmann, X. Lu, S.L. Carter, K. Cibulskis, S. Banerji, G. Getz, K.S. Park, D. Rauh, C. Grutter, M. Fischer, L. Pasqualucci, G. Wright, Z. Wainer, P. Russell, I. Petersen, Y. Chen, E. Stoelben, C. Ludwig, P. Schnabel, H. Hoffmann, T. Muley, M. Brockmann, W. Engel-Riedel, L.A. Muscarella, V.M. Fazio, H. Groen, W. Timens, H. Sietsma, E. Thunnissen, E. Smit, D.A. Heideman, P.J. Snijders, F. Cappuzzo, C. Ligorio, S. Damiani, J. Field, S. Solberg, O.T. Brustugun, M. Lund-Iversen, J. Sanger, J.H. Clement, A. Soltermann, H. Moch, W. Weder, B. Solomon, J.C. Soria, P. Validire, B. Besse, E. Brambilla, C. Brambilla, S. Lantuejoul, P. Lorimier, P.M. Schneider, M. Hallek, W. Pao, M. Meyerson, J. Sage, J. Shendure, R. Schneider, R. Buttner, J. Wolf, P. Nurnberg, S. Perner, L.C. Heukamp, P.K. Brindle, S. Haas, and R.K. Thomas, Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer, *Nat. Genet.* 44 (2012) 1104–1110.
- [42] C.M. Rudin, S. Durinck, E.W. Stawiski, J.T. Poirier, Z. Modrusan, D.S. Shames, E.A. Bergbower, Y. Guan, J. Shin, J. Guillory, C.S. Rivers, C.K. Foo, D. Bhatt, J. Stinson, F. Gnad, P.M. Haverly, R. Gentleman, S. Chaudhuri, V. Janakiramam, B.S. Jaiswal, C. Parikh, W. Yuan, Z. Zhang, H. Koeppen, T.D. Wu, H.M. Stern, R.L. Yauch, K.E. Huffman, D.D. Paskulin, P.B. Illei, M. Varella-Garcia, A.F. Gazdar, F.J. de Sauvage, R. Bourgon, J.D. Minna, M.V. Brock, S. Seshagiri, Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer, *Nat. Genet.* 44 (2012) 1111–1116.
- [43] M. Imielinski, A.H. Berger, P.S. Hammerman, B. Hernandez, T.J. Pugh, E. Hodis, J. Cho, J. Suh, M. Capelletti, A. Sivachenko, C. Sougnez, D. Auclair, M.S. Lawrence, P. Stojanov, K. Cibulskis, K. Choi, W.L. de, T. Sharifnia, A. Brooks, H. Greulich, S. Banerji, T. Zander, D. Seidel, F. Leenders, S. Ansen, C. Ludwig, W. Engel-Riedel, E. Stoelben, J. Wolf, C. Goparju, K. Thompson, W. Winckler, D. Kwiatkowski, B.E. Johnson, P.A. Janne, V.A. Miller, W. Pao, W.D. Travis, H.I. Pass, S.B. Gabriel, E.S. Lander, R.K. Thomas, L.A. Garraway, G. Getz, M. Meyerson, Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing, *Cell* 150 (2012) 1107–1120.
- [44] P.S. Hammerman, D.N. Hayes, M.D. Wilkerson, N. Schultz, R. Bose, A. Chu, E.A. Colloff, L. Cope, C.J. Creighton, G. Getz, J.G. Herman, B.E. Johnson, R. Kucherlapati, M. Ladanyi, C.A. Maher, G. Robertson, C. Sander, R. Shen, R. Sinha, A. Sivachenko, R.K. Thomas, W.D. Travis, M.S. Tsao, J.N. Weinstein, D.A. Wigle,

- S.B. Baylin, R. Govindan, M. Meyerson, Comprehensive genomic characterization of squamous cell lung cancers, *Nature* 489 (2012) 519–525.
- [45] Cancer Genome Atlas Research Network, Comprehensive molecular portraits of human breast tumours, *Nature* 490 (2012) 61–70.
- [46] S. Banerji, K. Cibulskis, C. Rangel-Escareno, K.K. Brown, S.L. Carter, A.M. Frederick, M.S. Lawrence, A.Y. Sivachenko, C. Sougnez, L. Zou, M.L. Cortes, J.C. Fernandez-Lopez, S. Peng, K.G. Ardlie, D. Auclair, V. Bautista-Pina, F. Duke, J. Francis, J. Jung, A. Maffuz-Aziz, R.C. Onofrio, M. Parkin, N.H. Pho, V. Quintanar-Jurado, A.H. Ramos, R. Rebollar-Vega, S. Rodriguez-Cuevas, S.L. Romero-Cordoba, S.E. Schumacher, N. Stransky, K.M. Thompson, L. Uribe-Figueroa, J. Baselga, R. Beroukchim, K. Polyak, D.C. Sgroi, A.L. Richardson, G. Jimenez-Sanchez, E.S. Lander, S.B. Gabriel, L.A. Garraway, T.R. Golub, J. Melendez-Zajgla, A. Toker, G. Getz, A. Hidalgo-Miranda, M. Meyerson, Sequence analysis of mutations and translocations across breast cancer subtypes, *Nature* 486 (2012) 405–409.
- [47] M.J. Ellis, L. Ding, D. Shen, J. Luo, V.J. Suman, J.W. Wallis, B.A. Van Tine, J. Hoog, R.J. Goiffon, T.C. Goldstein, S. Ng, L. Lin, R. Crowder, J. Snider, K. Ballman, J. Weber, K. Chen, D.C. Koboldt, C. Kandath, W.S. Schierding, J.F. McMichael, C.A. Miller, C. Lu, C.C. Harris, M.D. McLellan, M.C. Wendt, K. DeSchryver, D.C. Allred, L. Esserman, G. Unzeitig, J. Margenthaler, G.V. Babiera, P.K. Marcom, J.M. Guenther, M. Leitch, K. Hunt, J. Olson, Y. Tao, C.A. Maher, L.L. Fulton, R.S. Fulton, M. Harrison, B. Oberkfell, F. Du, R. Demeter, T.L. Vickery, A. Elhammali, H. Piwnica-Worms, S. McDonald, M. Watson, D.J. Dooling, D. Ota, L.W. Chang, R. Bose, T.J. Ley, D. Piwnica-Worms, J.M. Stuart, R.K. Wilson, E.R. Mardis, Whole-genome analysis informs breast cancer response to aromatase inhibition, *Nature* 486 (2012) 353–360.
- [48] P.J. Stephens, P.S. Tarpey, H. Davies, L.P. Van, C. Greenman, D.C. Wedge, S. Nik-Zainal, S. Martin, I. Varela, G.R. Bignell, L.R. Yates, E. Papaemmanuil, D. Beare, A. Butler, A. Cheverton, J. Gamble, J. Hinton, M. Jia, A. Jayakumar, D. Jones, C. Latimer, K.W. Lau, S. McLaren, D.J. McBride, A. Menzies, L. Mudie, K. Raine, R. Rad, M.S. Chapman, J. Teague, D. Easton, A. Langerod, M.T. Lee, C.Y. Shen, B.T. Tee, B.W. Huimin, A. Broeks, A.C. Vargus, G. Turashvili, J. Martens, A. Fatima, P. Miron, S.F. Chin, G. Thomas, S. Boyault, O. Mariani, S.R. Lakhani, d. van, V. van, V. J. Foekens, C. Desmedt, C. Sotiriou, A. Tutt, C. Caldas, J.S. Reis-Filho, S.A. Aparicio, A.V. Salomon, A.L. Borresen-Dale, A.L. Richardson, P.J. Campbell, P.A. Futreal, M.R. Stratton, The landscape of cancer genes and mutational processes in breast cancer, *Nature* 486 (2012) 400–404.
- [49] S.P. Shah, A. Roth, R. Goya, A. Oloumi, G. Ha, Y. Zhao, G. Turashvili, J. Ding, K. Tse, G. Haffari, A. Bashashati, L.M. Prentice, J. Khattra, A. Burleigh, D. Yap, V. Bernard, A. McPherson, K. Shumansky, A. Crisan, R. Giuliani, A. Heravi-Moussavi, J. Rosner, D. Lai, I. Birol, R. Varhol, A. Tam, N. Dhalla, T. Zeng, K. Ma, S.K. Chan, M. Griffith, A. Moradian, S.W. Cheng, G.B. Morin, P. Watson, K. Gelmon, S. Chia, S.F. Chin, C. Curtis, O.M. Rueda, P.D. Pharoah, S. Damaraju, J. Mackey, K. Hoon, T. Harkins, V. Tadigotla, M. Sigaroudinia, P. Gascard, T. Tlsty, J.F. Costello, I.M. Meyer, C.J. Eaves, W.W. Wasserman, S. Jones, D. Huntsman, M. Hirst, C. Caldas, M.A. Marra, S. Aparicio, The clonal and mutational evolution spectrum of primary triple-negative breast cancers, *Nature* 486 (2012) 395–399.
- [50] T.J. Pugh, S.D. Weeraratne, T.C. Archer, D.A. Pomeranz Krummel, D. Auclair, J. Bochicchio, M.O. Carneiro, S.L. Carter, K. Cibulskis, R.L. Erlich, H. Greulich, M.S. Lawrence, N.J. Lennon, A. McKenna, J. Meldrim, A.H. Ramos, M.G. Ross, C. Russ, E. Shefler, A. Sivachenko, B. Sogoloff, P. Stojanov, P. Tamayo, J.P. Mesirov, V. Amani, N. Teider, S. Sengupta, J.P. Francois, P.A. Northcott, M.D. Taylor, F. Yu, G.R. Crabtree, A.G. Kautzman, S.B. Gabriel, G. Getz, N. Jager, D.T. Jones, P. Lichter, S.M. Pfister, T.M. Roberts, M. Meyerson, S.L. Pomeroy, Y.J. Cho, Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations, *Nature* 488 (2012) 106–110.
- [51] J.S. Welch, T.J. Ley, D.C. Link, C.A. Miller, D.E. Larson, D.C. Koboldt, L.D. Wartman, T.L. Lamprecht, F. Liu, J. Xia, C. Kandath, R.S. Fulton, M.D. McLellan, D.J. Dooling, J.W. Wallis, K. Chen, C.C. Harris, H.K. Schmidt, J.M. Kalicki-Veizer, C. Lu, Q. Zhang, L. Lin, M.D. O'Laughlin, J.F. McMichael, K.D. Delehaunty, L.A. Fulton, V.J. Magrini, S.D. McGrath, R.T. Demeter, T.L. Vickery, J. Hundal, L.L. Cook, G.W. Swift, J.P. Reed, P.A. Aildredge, T.N. Wylie, J.R. Walker, M.A. Watson, S.E. Heath, W.D. Shannon, N. Varghese, R. Nagarajan, J.E. Payton, J.D. Baty, S. Kulkarni, J.M. Kilo, M.H. Tomasson, P. Westervelt, M.J. Walter, T.A. Graubert, J.F. DiPersio, L. Ding, E.R. Mardis, R.K. Wilson, The origin and evolution of mutations in acute myeloid leukemia, *Cell* 150 (2012) 264–278.
- [52] M.A. Chapman, M.S. Lawrence, J.J. Keats, K. Cibulskis, C. Sougnez, A.C. Schinzel, C.L. Harview, J.P. Brunet, G.J. Ahmann, M. Adli, K.C. Anderson, K.G. Ardlie, D. Auclair, A. Baker, P.L. Bergsagel, B.E. Bernstein, Y. Drier, R. Fonseca, S.B. Gabriel, C.C. Hofmeister, S. Jagannath, A.J. Jakubowiak, A. Krishnan, J. Levy, T. Liefeld, S. Lonial, S. Mahan, B. Mfuko, S. Monti, L.M. Perkins, R. Onofrio, T.J. Pugh, S.V. Rajkumar, A.H. Ramos, D.S. Siegel, A. Sivachenko, A.K. Stewart, S. Trudel, R. Vij, D. Voet, W. Winckler, T. Zimmerman, J. Carpten, J. Trent, W.C. Hahn, L.A. Garraway, M. Meyerson, E.S. Lander, G. Getz, T.R. Golub, Initial genome sequencing and analysis of multiple myeloma, *Nature* 471 (2011) 467–472.
- [53] V. Quesada, L. Conde, N. Villamor, G.R. Ordonez, P. Jares, L. Bassaganyas, A.J. Ramsay, S. Bea, M. Pinyol, A. Martinez-Trillos, M. Lopez-Guerra, D. Colomer, A. Navarro, T. Baumann, M. Aymeric, M. Rozman, J. Delgado, E. Gine, J.M. Hernandez, M. Gonzalez-Diaz, D.A. Puentes, G. Velasco, J.M. Freije, J.M. Tubio, R. Royo, J.L. Gelpi, M. Orozco, D.G. Pisano, J. Zamora, M. Vazquez, A. Valencia, H. Himmelbauer, M. Bayes, S. Heath, M. Gut, I. Gut, X. Estivill, A. Lopez-Guillermo, X.S. Puente, E. Campo, C. Lopez-Otin, Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia, *Nat. Genet.* 44 (2012) 47–52.
- [54] E. Hodis, I.R. Watson, G.V. Kryukov, S.T. Arold, M. Imielinski, J.P. Theurillat, E. Nickerson, D. Auclair, L. Li, C. Place, D. Dicara, A.H. Ramos, M.S. Lawrence, K. Cibulskis, A. Sivachenko, D. Voet, G. Saksena, N. Stransky, R.C. Onofrio, W. Winckler, K. Ardlie, N. Wagle, J. Wargo, K. Chong, D.L. Morton, K. Stemke-Hale, G. Chen, M. Noble, M. Meyerson, J.E. Ladbury, M.A. Davies, J.E. Gershenwald, S.N. Wagner, D.S. Hoon, D. Schadendorf, E.S. Lander, S.B. Gabriel, G. Getz, L.A. Garraway, L. Chin, A landscape of driver mutations in melanoma, *Cell* 150 (2012) 251–263.
- [55] M.F. Berger, E. Hodis, T.P. Heffernan, Y.L. Deribe, M.S. Lawrence, A. Protopopov, E. Ivanova, I.R. Watson, E. Nickerson, P. Ghosh, H. Zhang, R. Zeid, X. Ren, K. Cibulskis, A.Y. Sivachenko, N. Wagle, A. Sucker, C. Sougnez, R. Onofrio, L. Ambrogio, D. Auclair, T. Fennell, S.L. Carter, Y. Drier, P. Stojanov, M.A. Singer, D. Voet, R. Jing, G. Saksena, J. Barretina, A.H. Ramos, T.J. Pugh, N. Stransky, M. Parkin, W. Winckler, S. Mahan, K. Ardlie, J. Baldwin, J. Wargo, D. Schadendorf, M. Meyerson, S.B. Gabriel, T.R. Golub, S.N. Wagner, E.S. Lander, G. Getz, L. Chin, L.A. Garraway, Melanoma genome sequencing reveals frequent PREX2 mutations, *Nature* 485 (2012) 502–506.
- [56] C.E. Barbieri, S.C. Baca, M.S. Lawrence, F. Demichelis, M. Blattner, J.P. Theurillat, E.A. White, P. Stojanov, A.E. Van, N. Stransky, I.A. Asangani, S.S. Chae, G. Boysen, D. Auclair, R.C. Onofrio, K. Park, N. Kitabayashi, T.Y. MacDonald, K. Sheikh, T. Vuong, C. Guiducci, K. Cibulskis, A. Sivachenko, S.L. Carter, G. Saksena, D. Voet, W.M. Hussain, A.H. Ramos, W. Winckler, M.C. Redman, K. Ardlie, A.K. Tewari, J.M. Mosquera, N. Rupp, P.J. Wild, H. Moch, C. Morrissey, P.S. Nelson, P.W. Kantoff, S.B. Gabriel, T.R. Golub, M. Meyerson, E.S. Lander, G. Getz, M.A. Rubin, L.A. Garraway, Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer, *Nat. Genet.* 44 (2012) 685–689.
- [57] C.S. Grasso, Y.M. Wu, D.R. Robinson, X. Cao, S.M. Dhanasekaran, A.P. Khan, M.J. Quist, X. Jing, R.J. Lonigro, J.C. Brenner, I.A. Asangani, B. Ateeq, S.Y. Chun, J. Siddiqui, L. Sam, M. Anstett, R. Mehra, J.R. Prensner, N. Palanisamy, G.A. Ryslik, F. Vandin, B.J. Raphael, L.P. Kunju, D.R. Rhodes, K.J. Pienta, A.M. Chinnaiyan, S.A. Tomlins, The mutational landscape of lethal castration-resistant prostate cancer, *Nature* 487 (2012) 239–243.
- [58] A. Kumar, T.A. White, A.P. MacKenzie, N. Clegg, C. Lee, R.F. Dumpit, I. Coleman, S.B. Ng, S.J. Salipante, M.J. Rieder, D.A. Nickerson, E. Corey, P.H. Lange, C. Morrissey, R.L. Vessella, P.S. Nelson, J. Shendure, Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers, *Proc. Natl. Acad. Sci. USA* 108 (2011) 17087–17092.
- [59] Cancer Genome Atlas Research Network, Comprehensive molecular characterization of human colon and rectal cancer, *Nature* 487 (2012) 330–337.
- [60] A. Fujimoto, Y. Totoki, T. Abe, K.A. Boroevich, F. Hosoda, H.H. Nguyen, M. Aoki, N. Hosono, M. Kubo, F. Miya, Y. Arai, H. Takahashi, T. Shirakihara, M. Nagasaki, T. Shibuya, K. Nakano, K. Watanabe-Makino, H. Tanaka, H. Nakamura, J. Kusuda, H. Ojima, K. Shimada, T. Okusaka, M. Ueno, Y. Shigekawa, Y. Kawakami, K. Arihiro, H. Ohdan, K. Gotoh, O. Ishikawa, S. Ariizumi, M. Yamamoto, T. Yamada, K. Chayama, T. Kosuge, H. Yamaue, N. Kamatani, S. Miyano, H. Nakagawa, Y. Nakamura, T. Tsunoda, T. Shibata, H. Nakagawa, Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators, *Nat. Genet.* 44 (2012) 760–764.
- [61] C. Guichard, G. Amaddeo, S. Imbeaud, Y. Ladeiro, L. Pelletier, I.B. Maad, J. Calderaro, P. Bioulac-Sage, M. Letexier, F. Degos, B. Clement, C. Balabaud, E. Chevet, A. Laurent, G. Couchu, E. Letouze, F. Calvo, J. Zucman-Rossi, Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma, *Nat. Genet.* 44 (2012) 694–698.
- [62] J.G. Lohr, P. Stojanov, M.S. Lawrence, D. Auclair, B. Chapuy, C. Sougnez, P. Cruz-Gordillo, B. Knoechel, Y.W. Asmann, S.L. Slager, A.J. Novak, A. Dogan, S.M. Ansell, B.K. Link, L. Zou, J. Gould, G. Saksena, N. Stransky, C. Rangel-Escareno, J.C. Fernandez-Lopez, A. Hidalgo-Miranda, J. Melendez-Zajgla, E. Hernandez-Lemus, C. Schwarz, I. Imaz-Rosshandler, A.I. Ojesina, J. Jung, C.S. Pedamallu, E.S. Lander, T.M. Habermann, J.R. Cerhan, M.A. Shipp, G. Getz, T.R. Golub, Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing, *Proc. Natl. Acad. Sci. USA* 109 (2012) 3879–3884.
- [63] J. Schwartzentruber, A. Korshunov, X.Y. Liu, D.T. Jones, E. Pfaff, K. Jacob, D. Sturm, A.M. Fontebasso, D.A. Quang, M. Tonjes, V. Hovestadt, S. Albrecht, M. Kool, A. Nantel, C. Konermann, A. Lindroth, N. Jager, T. Rausch, M. Ryzhova, J.O. Korbel, T. Hielscher, P. Hauser, M. Garami, A. Klekner, L. Bognar, M. Ebinger, M.U. Schuhmann, W. Scheurlen, A. Pekrun, M.C. Fruhwald, W. Roggendorf, C. Kramm, M. Durken, J. Atkinson, P. Lepage, A. Muntzpetit, M. Zakrzewska, K. Zakrzewski, P.P. Liberski, Z. Dong, P. Siegel, A.E. Kholzik, M. Zapotka, A. Guha, D. Malkin, J. Felsberg, G. Reifemberger, D.A. von, K. Ichimura, V.P. Collins, H. Witt, T. Milde, O. Witt, C. Zhang, P. Castelo-Branco, P. Lichter, D. Faury, U. Tabori, C. Plass, J. Majewski, S.M. Pfister, N. Jabado, Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma, *Nature* 482 (2012) 226–231.
- [64] K. Wang, J. Kan, S.T. Yuen, S.T. Shi, K.M. Chu, S. Law, T.L. Chan, Z. Kan, A.S. Chan, W.Y. Tsui, S.P. Lee, S.L. Ho, A.K. Chan, G.H. Cheng, P.C. Roberts, P.A. Rejto, N.W. Gibson, D.J. Pocalyko, M. Mao, J. Xu, S.Y. Leung, Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer, *Nat. Genet.* 43 (2011) 1219–1223.
- [65] Cancer Genome Atlas Research Network, Integrated genomic analyses of ovarian carcinoma, *Nature* 474 (2011) 609–615.
- [66] N. Stransky, A.M. Egloff, A.D. Tward, A.D. Kostic, K. Cibulskis, A. Sivachenko, G.V. Kryukov, M.S. Lawrence, C. Sougnez, A. McKenna, E. Shefler, A.H. Ramos, P.

- Stojanov, S.L. Carter, D. Voet, M.L. Cortes, D. Auclair, M.F. Berger, G. Saksena, C. Guiducci, R.C. Onofrio, M. Parkin, M. Romkes, J.L. Weissfeld, R.R. Seethala, L. Wang, C. Rangel-Escareno, J.C. Fernandez-Lopez, A. Hidalgo-Miranda, J. Melendez-Zajgla, W. Winckler, K. Ardlie, S.B. Gabriel, M. Meyerson, E.S. Lander, G. Getz, T.R. Golub, L.A. Garraway, J.R. Grandis, The mutational landscape of head and neck squamous cell carcinoma, *Science* 333 (2011) 1157–1160.
- [67] N. Agrawal, M.J. Frederick, C.R. Pickering, C. Bettegowda, K. Chang, R.J. Li, C. Fakhry, T.X. Xie, J. Zhang, J. Wang, N. Zhang, A.K. El-Naggar, S.A. Jasser, J.N. Weinstein, L. Trevino, J.A. Drummond, D.M. Muzny, Y. Wu, L.D. Wood, R.H. Hruban, W.H. Westra, W.M. Koch, J.A. Califano, R.A. Gibbs, D. Sidransky, B. Vogelstein, V.E. Velculescu, N. Papadopoulos, D.A. Wheeler, K.W. Kinzler, J.N. Myers, Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1, *Science* 333 (2011) 1154–1157.
- [68] Y.-F. Guan, G.-R. Li, R.-J. Wang, Y.-T. Yi, L. Yang, D. Jiang, X.-P. Zhang, Y. Ping, Application of next-generation sequencing in clinical oncology to advance personalized treatment of cancer, *Chinese J. Cancer* 31 (2012) 463–470.