The Cancer Cell Map Initiative: Defining the Hallmark Networks of Cancer

Nevan J. Krogan,1,2,3,4,* Scott Lippman,5,6 David A. Agard,1,7 Alan Ashworth,4,8 and Trey Ideker5,6,*
1California Institute for Quantitative Biosciences (QB3)
2Department of Cellular and Molecular Pharmacology
University of California, San Francisco, San Francisco, CA 94143, USA
3J. David Gladstone Institutes, San Francisco, CA 94143, USA
4Helen Diller Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94143, USA
5Department of Medicine
6Moores Cancer Center
University of California, San Diego, San Diego, CA 92093, USA
7Department of Biochemistry and Biophysics
8Department of Medicine
University of California, San Francisco, San Francisco, CA 92093, USA
*Correspondence: nevan.krogan@ucsf.edu (N.J.K.), tideker@ucsd.edu (T.I.)
http://dx.doi.org/10.1016/j.molcel.2015.05.008

Progress in DNA sequencing has revealed the startling complexity of cancer genomes, which typically carry thousands of somatic mutations. However, it remains unclear which are the key driver mutations or dependencies in a given cancer and how these influence pathogenesis and response to therapy. Although tumors of similar types and clinical outcomes can have patterns of mutations that are strikingly different, it is becoming apparent that these mutations recurrently hijack the same hallmark molecular pathways and networks. For this reason, it is likely that successful interpretation of cancer genomes will require comprehensive knowledge of the molecular networks under selective pressure in oncogenesis. Here we announce the creation of a new effort, The Cancer Cell Map Initiative (CCMI), aimed at systematically detailing these complex interactions among cancer genes and how they differ between diseased and healthy states. We discuss recent progress that enables creation of these cancer cell maps across a range of tumor types and how they can be used to target networks disrupted in individual patients, significantly accelerating the development of precision medicine.

A Tsunami of Cancer Genome Information

Over the last few years, a number of global consortia, such as The Cancer Genome Atlas (TCGA) (Cancer Genome Atlas Research Network, 2008, 2011, 2012a, 2012b, 2012c, 2013a, 2013b, 2014a, 2014b) and the International Cancer Genome Consortium (Hudson et al., 2010; Jones et al., 2012), are seeking to create a comprehensive catalog of the genes that initiate or cause the progression of cancer. These studies involve full exome or genome sequencing of tumor samples, followed by comparison to normal tissue genomes from the same patients, to identify the point mutations, copy number aberrations, and gross genomic rearrangements that have been acquired exclusively in the tumor. Such “somatic” events are examined statistically over a population of cancer patients to identify genes that are altered more often than expected by chance, resulting in a list of approximately 250 frequently mutated genes thus far (Lawrence et al., 2013, 2014). In parallel, numerous genome-wide association studies (GWAS) have compared the normal or
“germline” genomes of cancer patients to those of healthy controls to identify inherited genetic variants that are associated with increased risk of cancer (Amundadottir et al., 2009; Flaherty et al., 2010; Girotti et al., 2015; Kuchenbaecker et al., 2015; Long et al., 2014; Pharoah et al., 2013; Turnbull et al., 2010; Vijai et al., 2015; Yeager et al., 2009). To date, GWAS has identified around 200 associations between cancer risk and specific genetic loci, such as single-nucleotide polymorphisms, haplotypes, and rare variants (Chang et al., 2014). Somatic and germ-line variants are then accumulated in data repositories, including the Catalogue of Somatic Mutations in Cancer (COSMIC) (Forbes et al., 2015) and the National Human Genome Research Institute’s GWAS Catalog (Welter et al., 2014), as are the cancer genomes themselves (Wilks et al., 2014). Some of the cancer projects encompass data sets that span not just genome sequence but a variety of other layers of complementary ‘omics information, including epigenetic changes, RNA profiles, and protein abundances. Several projects are also generating related ‘omics and drug sensitivity data across large panels of cancer cell lines, including the Cancer Cell Line Encyclopedia (Barretina et al., 2012), the Genomics of Drug Sensitivity in Cancer Resource (Yang et al., 2013), and Project Achilles (Nijhawan et al., 2012).

Cancer as a Pathway-Based Disease

For simple Mendelian disorders, genome sequencing has been able to identify specific loci that are directly responsible for a given disease state (Figure 1A, left). For example, some of the greatest successes in GWAS have been in characterizing the polygenic nature of type 2 diabetes, lipid levels, and obesity (Teslovich et al., 2010). In other cases, large-scale GWAS have illuminated individual loci that dominate the genetic contribution to a complex trait, such as the case of HIV Elite Controllers (Pereyra et al., 2010). However, these insights are relatively atypical of GWAS, and more prevalent are Manhattan plots with suggestive signals but which lack statistical significance (Figure 1A, right side) (Goldstein, 2009; Maher, 2008; Manolio et al., 2009; Visscher et al., 2012; Zhang et al., 2014). In cancer and other complex diseases, the power to detect important germine variants or somatic mutations has been hampered by the extreme genetic heterogeneity observed across the patient population (Gerlinger et al., 2012; Lawrence et al., 2013; Longo, 2012; Zhang et al., 2014). In particular, one of the most striking outcomes of the cancer sequencing projects is the discovery that each tumor genome is quite different from every other. Although many tumors harbor alterations to well-known cancer genes such as TP53 or PIK3CA, beyond these common events lies a “long tail” of ~20 to >1,000 genetic alterations, the majority of which are rare across the patient population. Heterogeneity has also been observed within different subpopulations of cells taken from a single patients’ tumor (Fisher et al., 2013; Lawrence et al., 2013; Zhang et al., 2014) complicating the analysis of tumor genomes as frequently only the dominant clone is analyzed. This lack of common genetic signal between or within patients limits the power of gene association or frequency-based methods to link individual genes with pathogenesis in cancer (Yaffe, 2013).
As a common explanation for the observed heterogeneity, it has been widely discussed that cancer is a disease that arises because of the action of hallmark cancer pathways (Hanahan and Weinberg, 2000, 2011; Vogelstein and Kinzler, 2004). Although any particular mutation or mutated gene may be a rare event when viewed independently, the key hypothesis is that various rare mutations found in different individuals will be found to converge on a smaller number of protein complexes, signaling cascades, or transcriptional regulatory circuits. This “pathway principle” is being applied with increasing success in recent interpretations of cancer genomics data sets (Cancer Genome Atlas Research Network, 2008; Ding et al., 2010; Perry et al., 2014). For example, a recent analysis of TCGA data identified 16 significantly mutated genetic networks that correspond to well-known cancer signaling pathways, as well as networks with less characterized roles in cancer, including cohesin and condensin (Leiserson et al., 2015), two related protein complexes required for segregating sister chromatids during cell division. Other efforts have shown that, within a hallmark cancer pathway, typically no more than one gene is mutated per tumor. This effect, which has been termed “mutual exclusivity,” possibly occurs because additional mutations in that same pathway provide no further selective advantage (Ciriello et al., 2012). Other sets of genes are rarely observed to be mutated together in the same tumor due to synthetic lethal interaction, in which the presence of the first mutation creates a specific vulnerability to the second, providing negative selective pressure against that combination during tumor evolution (Ashworth et al., 2011; Corcoran et al., 2013; Gross et al., 2014; Jerby-Aron et al., 2014; Morrison et al., 1997; Nijman and Friend, 2013; Scholl et al., 2009; Vizeacoumar et al., 2013). Thus, the key question is not whether cancer is a pathway or network-based disease—this hypothesis is familiar and increasingly well supported—but how best to apply this principle to the analysis and interpretation of cancer genomes.

**A Network-Mapping Strategy for Cancer**

Several groups have strongly argued that what is needed to interpret genomic data is the ability to place these data into biological context by mapping the identified sequence alterations onto the complexes and pathways in which they function (Califano et al., 2012; Ideker et al., 2011; Pujana et al., 2007; Sahni et al., 2015; Wang et al., 2010; Yaffe, 2013; Zuk et al., 2012). Instead of Manhattan plots testing individual genes for association, generation of analogous plots would be performed using networks (Figure 1B). To effectively enable this type of analysis, we believe it will be absolutely critical to initially define networks in “healthy” cells, unperturbed by specific mutations. Using cancer genomic information, the sets of genes that are linked to different disease states are being identified, and these should be targeted for systematic study both with and without cancer mutations (Figure 1C). While many databases exist to house information on pathways and complexes compiled from individual studies, and should indeed be exploited, there is great power in the systematic and unbiased collection of network data to accurately recapitulate the molecular wiring of the cell. Ultimately, the goal should be to generate “differential” maps (Bandyopadhyay et al., 2010; Ideker and Krogan, 2012), in which networks are compared in the presence versus absence of the specific mutants associated with different disease states. Only by comparing to the reference “wild-type” map will it be possible to truly understand how mutated genomes hijack and rewire pathways and complexes during the course of disease.

In model organisms, systematic integration of physical and genetic interaction data is a powerful way to reconstruct biological pathways (Bandyopadhyay et al., 2008; Beltrao et al., 2010; Beyer et al., 2007; Collins et al., 2007; Fischer et al., 2015; Gao et al., 2004; Hannum et al., 2009; Kelley and Ideker, 2005; Rognve et al., 2008; Schuldiner et al., 2005; Zhu et al., 2008). Mapping of physical networks, including protein-protein, protein-DNA, protein-RNA and DNA-DNA interactions, is used to systematically identify the physical modules in a system, including macromolecular protein complexes. In contrast, genetic interaction data, in which genes are perturbed in a pairwise fashion and the effects quantitatively assessed, provide the functional framework to place the physical modules into meaningful pathways. We argue that a similar dual approach should be utilized to define and characterize the hallmark pathways of cancer. Recent advances in genome engineering and interaction mapping in mammalian systems have made it possible to systematically physically and genetically interrogate sets of genes and proteins that have been linked to different disease states. For example, CRISPR/Cas9 technology permits facile labeling of genes with affinity tags and the engineering of specific point mutants into the genomes of mammalian somatic cells (Doudna and Charpentier, 2014) or induced pluripotent stem cells (iPS cells) (Yamanaka, 2009). Importantly, targeting specific point mutants using network biology approaches will be key in understanding how these mutations give rise to different diseases states (Graberg et al., 2013; Sahni et al., 2013). The CRISPR/Cas9 recombination technology helps to ensure normal expression of the targeted protein, once tagged, in subsequent proteomics analysis. Since iPS cells can give rise to other cell types in the body, one can argue that targeting such a cell line would be appropriate, at least initially. Using specific genetic and chemical manipulation, these iPS cell lines can then be differentiated into specific cell types (Amabile and Meissner, 2009), proteins purified in a temporal fashion (Bisson et al., 2011; Ideker and Krogan, 2012), and complexes analyzed via mass spectrometry (Figure 1B, bottom left). While analysis of isolated mutations in iPS cells should be carried out, it will also be imperative to study mutations in the cancerous lines in which they arise. Therefore, we anticipate that a complete Cancer Cell Map must ultimately incorporate data generated using “healthy” cells as well as appropriate tumor cell lines (Figure 1C). There is certain thinking that, even in a single patient, cancer arises from a heterogeneous group of cell types, each contributing a different set of mutations to the ultimate growth of a tumor (Fisher et al., 2013; Lawrence et al., 2013). If so, complexity grows, although it would still be of great value to have a reference map in even one cell type subjected to the set of mutations uncovered in cancer genomes.

Importantly, the identification of protein complexes will facilitate structural studies, including those involving recently developed cryo-EM technology (Li et al., 2013; Shen et al., 2015).
Similarly, sets of genes can be analyzed functionally using genetic interaction mapping techniques (Bassik et al., 2013; Laufer et al., 2013; Roguev et al., 2013), and the resulting data can be used to place complexes into pathways (Figure 1B, bottom right). Finally, a number of cancers arise in part by viral infection, including human papillomavirus involved in oropharyngeal and cervical cancers (Gross et al., 2014; Marur et al., 2010; Munoz et al., 2006) and hepatitis B and C involved in liver cancer (Leverero, 2006). The corresponding pathogenic viral proteins can be used as probes to generate physical and genetic interaction maps to help better interpret the cancer genomic data sets, since similar mechanisms may exist between molecular re-wiring induced by viruses and in mutated cancer genomes. We have recently shown that targeting viruses for systematic host-pathogen interaction studies is a powerful approach to help understand the underlying biology behind different pathogenic disease states (Davis et al., 2015; Jäger et al., 2012a, 2012b; Ramage et al., 2015).

From Descriptive to Predictive Networks
Armed with a collection of systematically generated network maps, we see several immediate applications. First, there is increasing evidence that prior knowledge of molecular networks can powerfully guide the search for disease genes (Ideker et al., 2011). A number of studies have recently demonstrated the power of network-guided GWAS for diverse diseases such as autism (O’Roak et al., 2012; Parkshak et al., 2013; Wilsey et al., 2013), multiple sclerosis (Baranzini et al., 2009; International Multiple Sclerosis Genetics Consortium, 2013), and motor neuron disorders (Novario et al., 2014), among others. These efforts are fueled by a growing number of bioinformatic approaches for integrating genetic data with molecular interaction networks or pathway databases (Bandyopadhyay et al., 2008; Collins et al., 2007; Hannum et al., 2009; Jia et al., 2011; Kelley and Ideker, 2005; Roguev et al., 2008; Wang et al., 2010). Pathway and network-guided GWAS in the cancer field are in their infancy (Fehringer et al., 2012; Koster et al., 2014; Tang...
Growing databases of tumor genomes are mined to identify genes and cell types in which alterations drive or predispose to development of cancer. Seeded by this information, systematic network mapping efforts lead to assembly of the “Cancer Cell Map,” providing a working scaffold of molecular interactions and the cell types and conditions under which they are active. New patient data are assessed by query against this resource, which translates alterations at the genetic and molecular level to reveal the impact these alterations have on the hallmark networks of cancer. Key interactions and network structures are explored for fine mapping of their basis in molecular structure or implications to biomolecular mechanism.

et al., 2014), although highly related progress is being made in identifying new cancer genes based on integration of somatic mutations and networks (Ciriello et al., 2012; Leiserson et al., 2015). Regardless, the concepts and methodology are general, and we anticipate that, as with other diseases, future cancer genomic studies will rely increasingly on network knowledge.

We anticipate that a second major use of network information in cancer will be translational, by providing a platform for integrating patient data to make biological and clinical predictions. In an increasing number of cases, network analyses are yielding better indicators of patient survival times or responses to therapy than traditional ‘black-box’ statistical regression or classification methods (Carro et al., 2010; Chuang et al., 2007, 2012; Dutkowski and Ideker, 2011; Hofree et al., 2013; Jerby-Avron et al., 2014; Taylor et al., 2009; Vaske et al., 2010). We recently introduced the method of network-based stratification (NBS), which allows for stratification of cancer into informative subtypes by clustering together patients who have somatic mutations that impact common regions of a reference molecular interaction network (Hofree et al., 2013). Several publicly available molecular interaction networks can be effective aids in this task, including HumanNet, a network of approximately 500,000 human protein interactions which integrates 21 types of ‘omics data, with each data type weighted according to how well it links proteins known to function together (Lee et al., 2011). NBS has been demonstrated in ovarian, uterine, and lung cancer (Hofree et al., 2013) and, more recently, thyroid cancer (Giordano and Cancer Genome Atlas Research Network, 2014). For each of these tissues, clear subtypes are identified that are predictive of clinical outcomes such as patient survival or response to therapy. For example, the network region for ovarian cancer subtype 1 (Figure 2A) clusters patients with somatic mutations in genes including the Fibroblast Growth Factor Receptor (FGFR) family and cytoskeletal signaling proteins such as ankyrin (e.g., ANK1). Patients with mutations in this network region have the most aggressive ovarian tumors as measured by survival, in contrast to patients whose mutations place them in other subtypes (Figure 2B). Strikingly, the mutations of patients in subtype 1 do not form a recognizable cluster in the absence of network information (Figure 2C), since these mutations largely affect different genes.

Moving forward, strong emphasis should be placed on new methodologies to identify which networks and pathways are affected by cancer mutations; to use network knowledge to predict the outcome of therapy in a given patient; and to predict novel molecular targets based on network information (Watson et al., 2013). For these purposes, one can leverage databases of curated mutations, clinical data and drug-gene interactions, such as DoCM, CIViC and DGIdb (http://docm.genome.wustl.edu, https://civic.genome.wustl.edu; Griffith et al., 2013). Such tasks will require a bioinformatic framework to integrate patient molecular and clinical data with biological networks, with the goal of making clinically relevant diagnoses and predictions about an individual based on their genomic information. By probing the available network knowledge as part of routine analysis of patient data, questions such as which cancer subtype a patient belongs to, whether their tumor will metastasize or not, and what drug should be chosen to treat them may well be answered.

Toward Creation of a Cancer Cell Map

As we have outlined, there exists a vast amount of sequence data from tumors associated with many different cancer types, and efforts are ongoing to extract mechanistic insight from this information. Given all of this progress, what is now needed is to develop knowledge of the molecular networks that drive oncogenesis, an effort that will ultimately impact translational research. Genomic data derived from tumor sequencing studies identifies key genes implicated in different cancer states. Integrated physical and genetic networks based on these factors (Figure 1) will help to put the mutations into clinical context (Figure 2) as well as to discover new disease genes as interacting partners become apparent (Figure 3). This is the goal of The Cancer Cell Map, a resource that can be used for cancer genome interpretation and to identify key complexes and pathways to be studied in greater mechanistic detail to get a deeper understanding about the biology underlying different cancer states.

Although such a framework is readily applicable not only to cancer but also to virtually any other disease, an initial focus on cancer is warranted for several reasons. First, cancer is well known to be one of the greatest causes of mortality in developed nations. Second, many organized efforts focused on cancer, including those from TCGA and ICGC, have resulted in the collection of many somatic tumor exomes, methylomes, microRNA and mRNA profiles, as well as many other ‘omics data sets, providing the most concerted and comprehensive data collection for any disease, information that can be leveraged in the CCMI. Finally, unlike many other complex diseases, cancer is an evolutionary process driven by somatic changes to normal cells, a fact that allows for powerful subtractive approaches, which reduce many genomic variants to a few somatic alterations. We believe that similar subtractive approaches could and should be applied to systematic network mapping efforts (Ideker and Krogan, 2012). Ultimately a collection of cell maps generated to help understand the biology behind many disease states would provide a resource to gain insight into data collected from new patients, and, depending on the pathways...
and complexes that are perturbed, select targeted therapies, the goal of precision medicine (Figure 3).

ACKNOWLEDGMENTS

We would like to thank Hannah Carter, Ryan Hernandez, Danielle Swaney, Minkyu Kim, Michael Shales, and Matan Horef for helpful comments and discussion as well as the Fred Luddy Family Foundation for the founding donation to the CCMI.

REFERENCES


BRAF mutant melanoma. Cancer Cell
'

Molecular Cell
Review


