Published in final edited form as:

Nat Rev Genet. 2017 July ; 18(7): 395–410. doi:10.1038/nrg.2017.8.

Functional variomics and network perturbation: connecting genotype to phenotype in cancer

[†] Correspondence should be addressed to Song Yi (Syi2@mdanderson.org) and Nidhi Sahni (Nsahni@mdanderson.org).

Competing interests statement

The authors declare no competing interests.

Further Information

SIFT: <http://sift.jcvi.org/>

SIFT4G: <http://sift-dna.org/sift4g>

MutSig/MutSigCV: <http://archive.broadinstitute.org/cancer/cga/mutsig>

MuSiC: <http://gmt.genome.wustl.edu/packages/genome-music/>

PolyPhen-2: <http://genetics.bwh.harvard.edu/pph2/>

STRUM: <http://zhanglab.ccmb.med.umich.edu/STRUM/>

MutationTaster: <http://www.mutationtaster.org/>

ActiveDriver: <http://www.baderlab.org/Software/ActiveDriver>

MSEA: <http://bioinfo.mc.vanderbilt.edu/MSEA/>

CanBind: <http://canbind.princeton.edu>

e-Driver: <https://github.com/eduardporta/e-Driver.git>

ANNOVAR: <http://annovar.openbioinformatics.org/en/latest/>

CADD: <http://cadd.gs.washington.edu/>

GWAVA: <http://www.sanger.ac.uk/science/tools/gwava>

deltaSVM: <http://www.beerlab.org/deltasvm/>

GWAS3D: <http://jjwanglab.org/gwas3d>

DeepSEA: <http://deepsea.princeton.edu/job/analysis/create/>

CanPredict: <http://www.canpredict.org/>

FunSeq2: <http://funseq2.gersteinlab.org/>

SuSPect: <http://www.sbg.bio.ic.ac.uk/~suspect/>

BeAtMuSiC: <http://babylone.ulb.ac.be/beatmusic>

Structure-PPI: <http://structureppi.bioinfo.cnio.es/Structure>

dSysMap: <http://dsysmap.irbbarcelona.org/>

MutaBind: <http://www.ncbi.nlm.nih.gov/projects/mutabind/>

is-rSNP: <http://www.genomics.csse.unimelb.edu.au/is-rSNP>

HaploReg: <http://archive.broadinstitute.org/mammals/haploreg/haploreg.php>

OncoCis: <http://149.171.80.192/OncoCis/>

BayesPI-BAR: <http://folk.uio.no/junbaiw/BayesPI-BAR/>

Patrocles: <http://www.patrocles.org/>

SomamiR/SomamiR 2.0: <http://compbio.uthsc.edu/SomamiR/>

PolymiRTS: <http://compbio.uthsc.edu/miRSNP/>

DriverNet: <http://compbio.bccrc.ca/software/drivernet/>

TieDIE: <https://sysbiowiki.soe.ucsc.edu/tiedie>

OncoIMPACT: <https://sourceforge.net/projects/oncoimpact/>

VarWalker: <http://bioinfo.mc.vanderbilt.edu/VarWalker>

HotNet/HotNet2: <http://compbio.cs.brown.edu/projects/hotnet2/>

NBS: <http://idekerlab.ucsd.edu/software/NBS/>

Subject categories

Biological sciences / Genetics / Clinical genetics / Disease genetics / Cancer genetics[URI /631/208/2489/144/68]

Biological sciences / Chemical biology / Networks and systems biology[URI /631/92/360]

Biological sciences / Systems biology / Regulatory networks[URI /631/553/2711]

Biological sciences / Genetics / Genome / Genetic variation[URI /631/208/726/649]

Biological sciences / Molecular biology / Proteomics / Protein–protein interaction networks[URI /631/337/475/2290]

Biological sciences / Computational biology and bioinformatics / Protein analysis / Protein sequence analyses[URI /631/114/663/2009]

Biological sciences / Biological techniques / Mass spectrometry[URI /631/1647/296]

Biological sciences / Biochemistry / Proteins / DNA-binding proteins[URI /631/45/612/1229]

Biological sciences / Genetics / CRISPR-Cas systems / CRISPR-Cas9 genome editing[URI /631/208/4041/3196]

Biological sciences / Genetics / Mutation[URI /631/208/737]

Song Yi^{#1,†}, Shengda Lin^{#2}, Yongsheng Li^{#1}, Wei Zhao¹, Gordon B. Mills¹, and Nidhi Sahni^{1,3,†}

¹Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

²Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA

³Graduate Program in Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, Houston, TX 77030, USA

[#] These authors contributed equally to this work.

Abstract

Proteins interact with other macromolecules in complex cellular networks for signal transduction and biological function. In cancer, genetic aberrations have been traditionally thought to disrupt the entire gene function. It has been increasingly appreciated that each mutation of a gene could have a subtle but unique effect on protein function or network rewiring, contributing to diverse phenotypic consequences across cancer patient populations. In this Review, we discuss the current understanding of cancer genetic variants including the broad spectrum of mutation classes and the wide range of mechanistic effects on gene function in the context of signaling networks. We highlight recent advances in computational and experimental strategies to study the diverse functional and phenotypic consequences of mutations at the base-pair resolution. Such information is critical for understanding the complex pleiotropic effect of cancer genes, and provides a possible link between genotype and phenotype in cancer.

With rapidly evolving next-generation sequencing technologies, there has been an explosion in human genotypic information, particularly for disease mutations associated with numerous types of cancer^{1,2} (FIG. 1a). However, the functional paths by which heterogeneous genotypic variants lead to diverse phenotypic consequences remain largely unresolved^{3,4}. More importantly, how the multiple genomic aberrations present in a single tumor integrate into the tumor phenotype — including response to therapy and patient outcomes — remains an overarching gap in knowledge in the field^{5,6}. It is therefore critical to identify the functional roles of distinct cancer mutations. A ‘one gene, one function, one disease’ model cannot be reconciled with the complexity that different mutations in the same gene often lead to markedly diverse phenotypes^{7,8}. It is now clear that genes and gene products do not function in isolation, but rather interact with each other in cellular networks^{9,10}. A systems-level understanding of how cancer mutations affect signaling networks is pivotal for interpreting the complex genotype-to-phenotype relationships in terms of tumor behavior and patient outcomes^{11,12}. This more sophisticated functional understanding of mutations is key for distinguishing drivers from non-pathogenic passengers, as well as enhancing clinical diagnostics, prognostics and therapeutics¹³.

Systematic high-throughput functional ‘variomics’ platforms to assess mutation-specific network perturbations are beginning to emerge^{11,12}. Here we review recent advances in the identification and functional characterization of genomic variants in human cancers, as part of an initiative towards creating a functional landscape of the cancer genomes. Integration of

systems genetics with signaling networks will be crucial for prioritizing cancer-causing variants, and to uncovering patient mutation-specific disease mechanisms and their resultant therapeutic liabilities^{4,6,13}. Together this effort represents a critical step towards personalized medicine.

In this Review, we first provide a brief overview of different classes of genomic aberrations that underlie cancer heterogeneity. We then discuss the importance of systems genetics and network biology for understanding the functional effects of various cancer mutations. To distinguish driver mutations from passenger mutations, we describe a toolkit of recent computational and bioinformatics algorithms to prioritize genomic mutations and interpret their potential functional impact. To reveal molecular underpinnings driving selection of particular mutations, we further describe emerging systems genetics technologies that enable the functional characterization of cancer variants in a high-throughput manner. Finally, with the integration of computational and experimental platforms, we aim to provide systems-level strategies to stratify cancer mutations at single nucleotide resolution, therefore bridging genotype to phenotype in cancer.

For brevity, we do not discuss genetic interaction networks or metabolic networks (reviewed in REFS^{3,14,15}), modeling of network dynamics and the analysis of network centrality (reviewed in REFS^{3,6,16}), or the functional platforms used to assess variants in non-coding regions (reviewed in REF¹⁷), as these have been extensively reviewed previously.

Genetic heterogeneity in cancer

Genomic instability propels the accumulation of mutations in cancer cells, and results in the rapid evolution of cancer genomes both in response to the microenvironmental stresses that occur during tumor evolution and the stresses induced by tumor therapy. Identifying and characterizing cancer driver mutations is among the most pressing need to understand the causality and progression of tumors, as well as the development of effective treatments tailored to specific cancer patients. A plethora of sequencing data from recent whole-genome and whole-exome sequencing projects has begun to elucidate the mutational landscape of most common cancers, and reveal the functional and structural elements of cancer genomes. Here in this Review, we refer to ‘genetic heterogeneity’ primarily as distinct mutations between tumors or across patients.

Somatic versus germline mutations.

The majority of cancer mutations are somatic. Approximately 90% of cancer genes show somatic mutations and 20% show germline mutations, with 10% showing both somatic and germline mutations¹⁸. Some somatic mutations, such as those in telomerase reverse transcriptase (*TERT*) and *TP53* (which encodes p53), frequently occur across cancer lineages, while others can be more tissue specific. Somatic aberrations show much more diverse patterns compared to germline variants, including complex genomic rearrangements, such as chromoplexy¹⁹, chromothripsis²⁰, and kataegis²¹. This is probably due to much reduced evolutionary constraints in somatic cells, as somatic mutations only need to be compatible with viability of the subset of cells in which they are found, whereas germline-inherited mutations will be present body-wide throughout development. The accumulation

of somatic mutations in cancer cells is pivotal in cancer progression. The ‘two-hit’ hypothesis originally postulated that tumors develop as a consequence of a second somatic mutation occurring upon the first inherited or somatic mutation²². It is now clear that more than two hits are needed for full emergence of the cancer phenotype in most tumors with statistical models placing this at between 2 and 6 driver aberrations. Cancer is commonly viewed as an evolutionary process of genetic instability and natural selection, driven by ongoing accumulation of somatic mutations^{23,24}. The continuous somatic evolution of cancers then contributes to genetic heterogeneity¹ and clonal expansion²⁵ during tumor progression.

Coding versus non-coding mutations.

Protein-coding regions of the genome comprise the ‘exome’, which represents <2% of the whole genome, but contains ~85% of known disease-related variants²⁶ (FIG. 1b). However, this is likely to be a technology bias as until recently high-throughput approaches to identify disease-related variants in non-coding regions have been limited. Whole-genome sequencing (WGS) has been applied for deep understanding of non-coding regulation in human cancer. By contrast, whole-exome sequencing (WES) has been widely used and proven to be a reliable and cost-effective approach to reveal the somatic cancer mutation landscape of protein-coding regions²⁷ (FIG. 1c). Non-coding regions contain functional modules that can elicit profound effects on expression profiles. *cis*-regulatory elements include promoters, enhancers, silencers and insulators (FIG. 1b). Mutations in these elements can drastically alter gene regulation. A potent example is *TERT*, which is reactivated in 80–90% of human cancers²⁸, often through recurrent mutations in its promoter region^{29,30}. These mutations create *de novo* binding sites for the GABP transcription factors^{31,32}. Non-coding RNAs (ncRNAs) are also frequent targets of genomic aberrations^{33–35}.

Identification of mutations in the coding genome can both reveal pathways that underlie cancer progression and identify therapeutic drug targets. Recently, The Cancer Genome Atlas (TCGA) pan-cancer analysis³⁶ has identified numerous cancer aberrations and helped to assign the abnormalities into physical complexes and pathways. Indeed, the aggregate of mutations in a complex such as the SWI/SNF complex or in a pathway such as the homologous recombination pathway serves to establish their importance in the tumorigenic process and further emphasize the need to characterize aberrations not as independent events but rather as part of functional machines.

Driver versus passenger aberrations.

Identification of cancer variants has been dominated by genotyping-by-sequencing. Indeed, over 3 million independent variants in coding regions have been identified by sequencing, with only a small subset of these being functionally annotated. Importantly, not all genetic alterations are relevant to tumor progression. The terms ‘driver’ and ‘passenger’ distinguish causal versus random events in cancers³⁷. A driver mutation provides a selective advantage to the tumor clone in its microenvironment at some point during its history, but is not necessarily required to sustain tumor growth throughout the evolution of the tumor. Driver aberrations can be derived from major events, such as chromosomal gains and losses,

chromosomal shattering and chromosomal chains³⁸. Alternatively, driver mutations can be derived from mutational hot-spots, for example V600E mutation in *BRAF*³⁹.

In cancer, there are multiple classes of mutations: missense mutations, frame-shift mutations, silent mutations, nonsense mutations, insertions or deletions, non-coding mutations, and others. The relative frequency of the classes of genomic aberrations vary markedly across cancer lineages, although overall, missense mutations are by far the most frequent and dominant class of coding aberrations in human cancers (FIG. 2ab). Kidney clear-cell carcinoma, glioblastoma multiforme, hepatocellular carcinoma, acute myeloid leukemia, colorectal carcinoma, and endometrial carcinoma with the exception of the serous-like subtype, are dominated by single nucleotide mutations, whereas almost all serous ovarian and breast carcinoma samples, a large fraction of lung and head and neck squamous cell carcinomas are dominated by copy-number variations^{40,41}.

The prevalence of random mutations, non-cancer tissue mixed in tumors, clonal heterogeneity, and ploidy variation makes it difficult to accurately ‘call’ mutations irrespective of whether the mutation is a driver or passenger. There are several strategies to improve mutation calling. First, sufficient sequencing depth is required. WGS is often performed with 30–60 fold coverage, WES is often performed with 100–150 fold coverage, and targeted sequencing of candidate gene panels is typically performed at 200–2000 fold coverage³⁸. Second, accurate estimation of the background mutation rate is essential for the discrimination of driver versus passenger mutations, as passenger mutations are expected to be present at similar frequencies to background mutation rates whereas driver mutations are expected to be present at higher frequencies due to the selective advantage they engender. Background mutation rates have traditionally been derived from synonymous mutation rates⁴², intronic and untranslated region (UTR) mutations⁴³. Despite lower depth in WGS or WES readouts compared to targeted gene panels, WGS and WES allow background mutation rates to be estimated with higher accuracy. Other genomic parameters have also been taken into account in modern algorithms, such as MuSiC⁴⁴ and MutSigCV⁴⁵. Together, it is likely that mutation rates are not uniform across genomes⁴⁶ and that regional estimates of mutation rates that take into account replication timing and other factors are needed.

A cancer network rewiring perspective

Systems genetics: from pathways to cancer networks.

Initial studies of the genetic heterogeneity of human cancers originated from pioneering WES projects in breast, pancreatic, colorectal and brain cancers^{47–50}. These studies showed that cancer genomes are highly complex with between 50–100 somatic alterations in each tumor. The Catalogue of Somatic Mutations in Cancer (COSMIC), which catalogues somatic mutation frequencies in tumors and tumor-derived cell lines, as of the latest Release v76 (February 2016) contained over 3.9 million unique coding mutations in over 5,000 human genes from over a million tumor samples⁵¹. Once identified, cancer genes and mutations need to be assigned to functional and regulatory pathways. Common sets of conserved proteins participate in ‘core’ functional pathways that mediate essential cellular processes in different cell types and tissues⁵². However, it remains unclear how these

conserved pathways function in different contexts to achieve signaling specificity, and how mutations in cancer cells disrupt signaling pathways and cellular functions.

Although distinguishing causal disease variants from non-pathogenic polymorphisms is often insufficient to establish mechanisms or predict phenotypic outcomes, identifying causal mutations remains a key, but challenging, step for functional interpretation of cancer genomes⁵³. Classical gene knockout or knockdown approaches cannot always resolve the diverse biological functions mediated by different mutations of the same gene⁵⁴. Therefore, understanding how specific variants affect molecular interaction networks is critical for interpreting complex genotype-to-phenotype relationships in cancer^{3,14,54}.

Effect of genetic mutations on cancer networks.

Protein products of mutated cancer genes do not function in isolation, but are part of highly interconnected cellular networks³, which are often depicted as nodes (molecules) and edges (interactions)^{3,6,55} (FIG. 2c). In interaction networks, a genetic mutation can lead to either a complete gene knockout-like behavior, as loss of all of its interactions in the network, or alternatively, as interaction perturbation ('edgetic'), leading to the loss or gain of specific interactions¹² (FIG. 3a). Edgetic mutations tend to be located at interaction interfaces with protein partners. For example, the edgetic mutations R24C and R24H in cyclin-dependent kinase 4 (CDK4), which are associated with melanoma, reside at the protein interaction interface with the partner CDK inhibitor 2C (CDKN2C; also known as p18^{INK4C}) (FIG. 3b). Similarly, the edgetic mutation F194S in fructose-bisphosphatase 1 (FBP1) is located at the interaction interface (FIG. 3c).

Prioritization and comprehensive understanding of driver mutations requires the integration of systematic large-scale experimental approaches with computational algorithms. In the next two sections, we will cover these aspects in detail.

Computational prediction of drivers

To predict driver mutations, computational systems biology has established and implemented modeling algorithms based on existing big datasets in cancer. Here, we summarize recent computational methodologies for characterizing the function of cancer mutations from the network perspective and classify them into three main categories: node-level, edge-level, and subnetwork-level predictions (FIG. 4, TABLE 1 and Supplementary information S1 (table)).

Node-centered effects of cancer mutations

Sequence features.—One of the most commonly used methods for estimating the functional effect of mutations is sequence comparison based on multiple alignment. These methods assume that amino acid substitutions in highly conserved positions are deleterious (for example, SIFT^{56,57} and SIFT4G⁵⁸). Some algorithms also incorporate sequence-based data with clinical data for inferring the relationships among mutations and the affected genes⁴⁴. In addition, drivers can be identified by finding genes that harbor significantly more mutations than expected by chance^{59,60}, such as MutSig⁴⁵ and MSEA⁶¹. Together, although these methods are informative in driver prioritization, their accuracy may be influenced by

empirically observed local mutation frequencies. Fortunately, the available large-scale WES or WGS datasets from several human genome projects, such as the 100,000 genomes project⁶² and the Icelandic genome project⁶³, may provide valuable information for building background mutation models. However, while these algorithms have reasonable predictive value for loss-of-function aberrations, they are not very accurate in predicting gain of function or edgetic aberrations.

Structural features.—Computational tools have been designed by integrating structural features with sequence information (for example, PolyPhen-2⁶⁴ and STRUM⁶⁵). Some methods rely on the evidence that many driver mutations recurrently occur in specific structural regions of proteins (e.g. protein domains and disordered regions)^{2,66–69} or disrupt the active sites (e.g. phosphorylation sites)^{70–72}. Recent methods map genetic mutations onto protein three-dimensional structures to evaluate the functional impact of mutations at high resolution^{73–75}. In addition, some algorithms integrate multiple evolutionary and structural features to evaluate the disease-causing potential of mutations, such as CanDrA⁷⁶ and MutationTaster⁷⁷. Last but not least, other methods prioritize driver mutations based on their location at the structural binding sites for small molecules (for example, CanBind⁷⁸ and SGDriver⁷⁹). Notably, most of these approaches predict mutational effects on the function of coding genes.

Regulatory features.—The Encyclopedia of DNA Elements (ENCODE) project⁸⁰ has provided a comprehensive map of regulatory elements by advanced techniques such as chromatin immunoprecipitation followed by sequencing (ChIP-seq), DNase-seq, and chromosome conformation capture. A number of computational methods to investigate the regulatory effects of cancer mutations have been proposed based on various regulatory features in ENCODE (for example, ANNOVAR^{81,82}). To score the deleterious consequences of genetic variants, some tools integrate a wide range of annotations (including genomic and epigenomic features) into one metric (for example, CADD⁸³, GWAVA⁸⁴, FitCons⁸⁵ and deltaSVM⁸⁶). Although these methods help prioritize driver mutations, they often neglect the chromatin structural context of these regulatory regions. To overcome this problem, other algorithms were developed for predicting driver variants through the integration of chromatin effects⁸⁷ and high-dimensional regulatory interactions⁸⁸. Together, these predictive methods have established a useful toolkit to explore the function of cancer mutations in regulatory regions, and their effect on the interactions with target genes.

Significantly mutated subnetworks or pathways in cancer

Although the above-mentioned computational methods have reached a reasonable level of accuracy for predicting loss-of-function aberrations, multiple lines of evidence have shown that integration of gene ontology and network features is critical in providing a holistic measure of the functional consequence of a cancer mutation. Some algorithms assess the functional impact of cancer mutations taking into account the observation that genes with distinct ontology terms possess different baseline tolerance for deleterious mutations⁸⁹. To further enhance the predictive power, other methods prioritize cancer-causing mutations by integrating sequence and structural features with gene ontology similarity (for example, CanPredict⁹⁰). Recently, the rapid accumulation of protein interaction network data has

provided a new basis for studying the topological features of cancer genes and mutations in cellular networks. It has been shown that cancer genes tend to possess high topological centrality, even higher than that found in essential genes. Several predictive algorithms combine network centrality with large-scale genomic resources for prioritizing variants in cancer (for example, SuSPect⁹¹ and FunSeq2⁹²). It has been demonstrated that network centrality helps to discriminate disease-associated versus tolerated mutations.

An important observation from analyses of the landscape of cancer mutations is that different tumors or cancer patients exhibit distinctly different mutational profiles. Furthermore, mutated genes tend to fall into a limited number of recurrently mutated subnetworks or pathways. This observation has stimulated systems-level approaches for detecting possible driver mutations, and integrative analyses to identify significantly altered pathways. Using network approaches, several algorithms prioritize mutations in cancer based on their effects on transcriptional output (for example, DIGGIT⁹³) or their links to dysregulated genes from gene expression data (for example, DriverNet⁹⁴, TieDIE⁹⁵ and OncoIMPACT⁹⁶). These algorithms are informative in identifying network modules that are related to downstream transcriptional changes induced by cancer mutations. In addition, some methods identify cancer or subtype related subnetworks by diffusing cancer mutations throughout a network based on network propagation process (for example, VarWalker⁹⁷, HotNet⁹⁸, NBS¹³ and HotNet2⁹⁹).

Edgetic effects of cancer mutations

Next, we discuss the computational methods to functionally characterize the edgetic effects of cancer mutations, especially in the context of protein–protein, transcription factor–gene and microRNA–gene interaction networks.

Protein–protein interaction context.

The prediction of the impact of a cancer mutation on protein–protein binding can be used to identify driver mutations. Some methods predict deleterious mutations based on the mutation-induced changes in binding free energy (BeAtMuSiC¹⁰⁰) or in their 3D protein complex context (Structure-PPI¹⁰¹). Other methods also evaluate the effects of mutations on protein interactions based on force fields and statistical potentials and fast side-chain optimization algorithms (for example, MutaBind¹⁰²). More globally, several algorithms (for example, dSysMap¹⁰³) predict drivers by mapping missense mutations onto the structurally annotated human interactome from Interactome3D¹⁰⁴, which is a valuable resource for exploring the edgetic role of disease mutations. A disadvantage of these methods is that they rely on known three-dimensional structures that are only available for a small proportion of proteins. By design, most studies of edgetic mutations have focused on loss of interaction; however, it is likely that cancer mutations could also result in a gain of interaction¹⁰⁵. These edgetic effects are as yet poorly predicted by current algorithms and require both development of new algorithms and iterative improvement of these algorithms with experimental data^{71,72}.

Gene regulatory context.

Edgetic modeling of cancer mutations is not limited to protein interactions but can be applied to transcription regulatory networks in different contexts^{106,107} and in noncoding regions^{11,12}. Many mutations identified by genome-wide association studies (GWAS) are likely to be regulatory single nucleotide polymorphisms (SNPs) that affect the ability of a transcription factor to bind DNA. Based on this hypothesis, some algorithms score mutant alleles with a position weight matrix (PWM) to detect disruptive transcription factor mutations (for example, is-rSNP¹⁰⁸). Additional tools annotate cancer drivers through calculating the change of a binding site caused by genetic mutations (for example, HaploReg¹⁰⁹ and OncoCis¹¹⁰). Moreover, biophysical modeling of protein–DNA interactions helps predict SNPs that cause significant changes in the binding affinity of transcription factors (for example, BayesPI-BAR¹¹¹). Finally, the complex miRNA–gene regulatory networks have been shown to control many key cellular processes that are dysregulated in cancers^{112,113}. Several databases have been constructed for compiling mutations that are predicted to perturb miRNA-mediated gene regulation, such as Patrocles¹¹⁴, SomamiR¹¹⁵ and PolymiRTS^{116–118}. Together, these available tools are invaluable in predicting a large number of cancer-associated regulatory mutations in signaling networks.

Systems-level experimental platforms

Functional analysis of cancer genes and mutations is key to understanding tumorigenic mechanisms and to developing therapeutic methods. Advances in large-scale experimental platforms and screens have revolutionized our ability to study cancer mutations, and have begun to reveal the functional networks of cancer mutations. Here we focus on recent advances on functionally characterizing coding mutations at large scale.

Transcriptome profiles altered by mutations.

Transcriptome profiling has been extensively used in the past decade for functional genomics^{40,119}. In human cancer, gene expression profiles in the tumors are often compared with that in the matched controls. RNA sequencing (RNA-seq) is one of the most common approaches for transcriptomic studies. In RNA-seq, total RNA is extracted from mutant or control samples, followed by reverse transcription to generate a cDNA library (FIG. 5a). After adding adaptors, RNA-seq samples can then be processed with next-generation sequencing. Computational algorithms are available to facilitate downstream RNA-seq data analysis. Recently, another transcriptomic platform has emerged: the library of integrated network-based cellular signatures (LINCS) L1000¹²⁰. LINCS L1000 can profile gene expression changes following genetic perturbation (mutations) of cell lines at high throughput. L1000 detects transcript abundance with optically addressed microspheres and a flow cytometric system (FIG. 5b). As a result, L1000 can directly measure a reduced representation of the transcriptome caused by a genetic mutation. Taken together, these techniques are robust in their assessment of global RNA expression levels for a given genetic background.

Proteome profiles altered by mutations.

To monitor protein expression levels in mutants at large scale, antibody-based platforms for targeted or global quantification of protein expression have been developed. Reverse-phase protein array (RPPA) technology (FIG. 5c) is a type of protein microarray that uses antibodies to detect relative expression levels of proteins in tissue or cell lysates from hundreds of samples simultaneously. This has been applied to a large number of mutation-bearing tumor samples from patients with cancer. A stringent antibody validation procedure must be in place to ensure the sensitivity, specificity and robustness of the platform. In addition, RPPA allows protein profiling with a small amount of sample and is a cost-effective technology. At present, most RPPA datasets usually include 150–300 antibodies that measure total proteins or specific modifications, such as phosphorylation, cleavage and fatty acid alteration.

Due to these technical advantages, a number of studies have used the RPPA platform to analyze the expression of proteins involved in cell cycle progression, apoptosis, signaling network activities and other key pathways that are associated with specific mutation variants. Furthermore, RPPA has proved to be an efficient approach to assess functions of rare mutations in a high-throughput manner and to identify driver mutations. A recent study¹²¹ characterizing *PIK3CA* mutations using RPPA demonstrated that even though the impact of its hotspot mutations has been well-established, differential oncogenic activity and variant-specific activation of PI3K signaling and other pathways, such as the MAPK pathway, have been observed across lower frequency mutations. RPPA has also been used to characterize mutations in *PIK3RI*, demonstrating that mutations can have marked edgetic effects through interrupting protein–protein interactions (PPIs) and also through gaining new interaction partners^{122–124}. These studies suggest mutation-specific targeting as a potentially more efficient approach in precision medicine.

Affinity purification coupled with mass spectrometry (AP-MS) allows the detection of protein expression and PPIs in near-physiological conditions. After affinity purification using an antibody against the bait protein, MS is then used to provide global and targeted profiling of protein expression (or modification) (Fig. 5d). Following normalization, the peak intensity patterns are analyzed and compared across multiple samples¹²⁵. To study interaction changes by cancer mutations, AP-MS has been applied to characterize changes in PPIs by melanoma-associated mutations in human CDK4¹²⁵. A variety of quantitative techniques have been applied to MS-based proteomic analysis, including label-free, metabolic labeling (e.g. stable isotope labeling with amino acids in cell culture (SILAC)) and chemical labeling (e.g. isobaric tag for relative and absolute quantitation (iTRAQ) or tandem mass tag (TMT)). Recent MS-based studies using iTRAQ labeling have demonstrated the ability to identify 8,000–11,000 proteins and 25,000 phosphosites per tumor on average^{126,127}. However, these approaches require a large amount of input material, remain time consuming and are costly. Together, although restricted by the number of available antibodies, RPPA is a robust approach to evaluate protein expression levels under different conditions or in distinct cell types. In contrast, AP-MS provides a complementary assessment of protein expression, which is global but less specific.

Protein–protein interaction changes induced by mutations.

Genetic mutations can impair protein interactions. Mechanistic understanding of cancer-associated mutations requires finding the molecular interactions and biochemical activities that these mutations perturb. For example, the cancer-associated C305F mutation in the zinc finger domain of MDM2 causes the loss of its binding to ribosomal proteins. This interaction perturbation disrupts the ribosomal stress response, contributing to tumorigenesis¹²⁸. In addition, the missense mutation R172H of p53 leads to a gain of interaction with the tumor suppressor DAB2IP, which inhibits DAB2IP function and increases the invasive behavior in cancer cells¹⁰⁵. Although some cancer mutations have been characterized, the functional mechanism behind most variants remains elusive^{4,7}. Network approaches using interactome maps have been successful for highlighting candidate cancer genes and disease modifier genes^{9,129}; however, the impact of most causal variants on interaction networks remains largely unknown.

Recently, several large-scale functional variomics and proteomics platforms have been applied to profile mutation-induced changes of molecular interactions, PPIs especially, relative to their wild-type counterparts. The high-throughput Gateway-compatible enhanced yeast two-hybrid¹³⁰ (HT-eY2H) system (Fig. 5e) and the protein fragment complementation assay (PCA)¹³¹ (Fig. 5f) have been implemented to detect PPI alterations. In these systems, the pair of protein partners is each fused to a fragment of a transcription factor or an enzyme, which is stable but inactive in isolation, whereas PPIs reconstitute the transcription factor or enzyme function. A recent investigation of genetic variant-specific effects on PPIs at large scale across diverse human diseases including cancer identified that, in comparison to non-disease polymorphisms, disease mutations were more likely to associate with interaction perturbations¹².

Protein–DNA interaction changes by mutations.

WGS has revealed abundant genetic variation affecting not only coding sequences, but also non-coding regulatory elements. For example, many point mutations in the transcription factor RUNX1 cause defective DNA binding, resulting in a familial platelet disorder predisposed to acute myeloid leukemia¹³². Frequent mutations detected in the *TERT* promoter create *de novo* binding motifs for ETS transcription factors, and play a central role in cancer-specific telomerase activation^{31,32}. Although protein–DNA interactions have been characterized in a few cases, it remains unclear how the majority of transcription factor mutations or non-coding DNA mutations affect their interactions.

ChIP-seq assays have been extensively used to map genome-wide transcription factor occupancy profiles (Fig. 5g), such as the ENCODE Project^{133,134}. In addition, several systems biology platforms have emerged to study protein–DNA interactions at large scale^{11,107,135,136}, such as enhanced yeast one-hybrid (eY1H) and protein-binding microarrays (PBMs). In the eY1H assay, a putative regulatory DNA sequence is used as bait to search for transcription factors that bind to that DNA sequence in yeast cells¹³⁷. A reporter gene is often placed downstream of the DNA sequence to assess protein–DNA interactions (FIG. 5h). Using these assays, a systematic study on the binding of ~1,000 human transcription factors to a large number of enhancer mutations found widespread

protein–DNA interaction perturbations in disease, which correlate well with target gene expression changes¹¹. Finally, large-scale TF-binding activity could be evaluated using PBMs¹⁰⁷ (FIG. 5i). A recent systematic study investigated TF variants for their DNA-binding affinity using PBMs, and identified that individuals with distinct mutations have unique TF DNA-binding profiles, which may contribute to phenotypic variation¹⁰⁷.

Pleiotropic mutational effects and integrative analysis.

In cancer, mutations occur in the context of genomic, transcriptomic and/or epigenetic aberrations. To gain a systems-level understanding of the functional effects of mutations, an integrative analysis of multi-omics datasets (such as gene expression, DNA copy-number, and DNA methylation) is critical. Several approaches have been proposed to address this direction. XSeq¹³⁸ analyzes the impact of somatic mutations by incorporating gene expression, patient mutations, and a gene interaction network. PARADIGM-SHIFT¹³⁹ is another example that infers mutated gene activity from gene expression and copy number in the context of genetic pathways. Integration of proteomics analysis with genomic data has enabled the detection of global proteomic patterns associated with potential driver genetic lesions. For example, a study of lung adenocarcinoma cell lines integrated differential protein expression data with distinct p53 mutational status, and identified an enrichment of key functional pathways, including epithelial adhesion, immune and stromal cells, and mitochondrial function¹⁴⁰. Given that genomic mutations often act in a cell type-specific and condition-dependent manner, such integrative modeling is more likely to resolve functional effects of mutations by controlling for other changes.

Functional validation by CRISPR in cancer.

CRISPR has emerged as a powerful and flexible tool^{141–144} to systematically interrogate cancer genomes. Together with other technologies, CRISPR has produced valuable data on the identification of new cancer genes as well as on the functional consequence of driver mutations. In the most widely used approach for CRISPR-based genome editing, a CRISPR-associated (Cas) nuclease, usually *Staphylococcus pyogenes* Cas9, is guided to a genomic target site by single-guide RNAs (sgRNA), where it creates DNA double-strand breaks (DSBs)¹⁴⁵. DSBs are typically repaired by non-homologous end-joining (NHEJ), leaving a random sequence scar of a small insertion or deletion (indel) that can inactivate the targeted gene of interest (Fig. 5j)^{146,147}. Current genome-wide CRISPR screen libraries contain $7 \times 10^4 - 2 \times 10^5$ sgRNAs, with 3–12 sgRNAs for each gene^{148–155}. Current CRISPR screens are useful for revealing gene-level information, but they lack the resolution to distinguish different mutations within a gene.

To precisely model specific cancer mutations by CRISPR, a DNA template (single-stranded or double-stranded) of homology is provided to convert the site of DSB to a desirable sequence, in a process known as homology-directed repair (HDR) (Fig. 5j). However, HDR is relatively inefficient compared to NHEJ, and can be further corrupted by indels. Some recent efforts have been made to improve the applicability of HDR in specific mutation editing^{156–159}. Synchronizing the expression of Cas9 with cell-cycle progression¹⁶⁰, or treating cells with two small molecules, L755507 and Brefeldin A¹⁶¹, could improve the HDR efficiency. The discovery of the smaller-size Cas9 from *Staphylococcus aureus* allows

the packaging of Cas9 and sgRNA expression constructs into the highly versatile adeno-associated virus (AAV) delivery vehicle¹⁶², enabling efficient HDR in specific organs such as the liver¹⁶³.

Recently, direct base change has been achieved by fusing nuclease-dead Cas9 (dCas9) with a cytidine deaminase^{41,164–166}, such as the activation-induced cytidine deaminase (AID), and rat-origin APOBEC1. These fusion proteins drive somatic hypermutation in locations close to the CRISPR target, thus creating genetic variants at a defined genomic locus without creating DNA breaks. The efficiency of base-editing can be further improved by a second fusion to a bacteriophage uracil glycosylase inhibitor (UGI) and restoration of the nicking activity of dCas9, resulting in a third-generation base editor (BE3, APOBEC–XTEN–dCas9(A840H)–UGI) that mediates C→T conversion with up to 37% efficiency¹⁶⁶. Taken together, the ability to induce specific mutations by CRISPR does provide an emerging and powerful tool for the analysis of functional consequences of candidate aberrations in genes in a low-throughput manner. However, despite constant advances made in the CRISPR technology, it remains inefficient for precise editing of genome sequences, making it challenging to apply to modeling the greater complexity edgetic effects of cancer mutations.

Conclusion

In this Review, we have summarized computational and high-throughput experimental strategies to systematically characterize cancer genetic mutations in the context of molecular interaction networks at base-pair resolution. Recent advances in systems biology and next-generation sequencing have facilitated the development of functional variomics platforms to evaluate the impact of a large number of disease mutations. Widespread protein–protein and protein–DNA interaction perturbations have been identified across various types of human diseases including cancer. It has been demonstrated that different mutations in the same gene frequently result in different interaction perturbation profiles. Altogether, interaction perturbation profiling of disease mutations provides a paradigm for dissecting heterogeneous genetic variants across cancer patient populations that is sorely needed given the large number of uncharacterized patient mutations and their potential impact on cancer phenotype and therapeutic liabilities.

During tumor evolution, mutations arise and accumulate in response to stress signals from its microenvironment or tumor therapy. During this process, a driver mutation occurs and confers the tumor a selective advantage. While other passenger mutations do occur, they do not provide any growth advantage. Genetic heterogeneity develops across diverse tumor populations over time. Numerous computational algorithms and tools have been developed to prioritize cancer mutations, based on different node- or edge-level functional properties. Although integrative computational analyses achieve some levels of accuracy in predicting disease-causing candidate mutations, a substantial fraction of the top hits should still be experimentally validated. The rapidly increasing functional annotation of cancer-specific mutations from variomics platforms offers the opportunity to iteratively improve the computational predictive tools based on high-quality test data. Furthermore, computational algorithms are often limited in predicting interaction perturbations or gain of function, which are common mutational effects on cancer signaling networks.

Systematic characterization of cancer variants for their effect on interaction networks is critical to a systems-level understanding of genetic heterogeneity. So far we have focused on heterogeneity across tumor samples, however, intra-tumor heterogeneity also exists. A tumor is made up of many cell types, each of which would have its own set of mutations and underlying networks. Interaction profiles of cancer mutations provide a fundamental link between genotype and phenotype. Network perturbation by mutations allows for grouping of distinct genotypes that share common effects on interaction profiles underlying a particular phenotype. In addition, the identified perturbed interaction partners allow us to uncover specific targets that are impaired in a mutation-specific context, which may in turn suggest therapeutic biomarkers guiding potential personalized precision medicine. However, the phenotypic diversity of cell types remains an important challenge for network biology: available network databases might not faithfully represent the particular cancer cell types to be investigated; furthermore, cell type composition and infiltration by stromal and immune-system cells might impact network wiring.

One needs to take into account cell-type specificity and tumor microenvironment to obtain a comprehensive understanding of cancer mutations. Integration of context-dependent computational resources and improved algorithms would help to deconvolute the functional effects of mutations in a cell type-specific fashion. In addition, emerging technologies such as single-cell experimental approaches could help further stratify mutational effects. To construct higher-resolution functional networks, it would be essential to incorporate multiple properties of genetic mutations, including gene expression, protein folding and structure, protein–protein and protein–DNA interactions and beyond.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

N.S. would like to acknowledge the following grants: the Cancer Prevention and Research Institute of Texas (CPRIT) New Investigator Grant RR160021, the University of Texas System Rising STARS award, the NIH/NCI grants P30CA016672, U54HG008100 and U01CA168394, and the University Center Foundation via the Institutional Research Grant program at the University of Texas MD Anderson Cancer Center.

Author details (for online version)

Song Yi: Song Yi is a faculty member of Systems Biology at the University of Texas MD Anderson Cancer Center (UTMDACC), Houston, USA. He received his B.S. from Peking University, China, and his Ph.D. with high distinction from the University of Iowa, Iowa City, USA. He did his postdoctoral training at Harvard Medical School and Dana-Farber Cancer Institute, Boston, Massachusetts, USA. He also did further training in Genomic Data Science from Johns Hopkins University, USA. His research is primarily focused on a systematic and quantitative understanding of molecular signaling networks in human health and disease, taking advantage of a combination of experimental and computational approaches.

Homepage: http://faculty.mdanderson.org/Song_Yi/Default.asp

Shengda Lin: Shengda Lin is a Research Associate at Stanford University, California, USA. His research interests include aging, cancer and stem cells. He integrates cellular, genetic and genomic tools to study the clonal dynamics in mammalian organs. He received his Ph.D. from the University of Iowa, Iowa City, USA, where he studied embryonic development and WNT signaling pathways.

Yongsheng Li: Yongsheng Li is a postdoctoral fellow at the Department of Systems Biology, the University of Texas MD Anderson Cancer Center, Houston, USA. His main research interest is the application of computational methods to human cancer for better understanding complex disease etiology, particularly focused on genomic mutations, miRNA and lncRNA regulations. He received his Ph.D. in Bioinformatics from the Harbin Medical University, China.

Wei Zhao: Wei Zhao is a postdoctoral fellow at the Department of Systems Biology at the University of Texas MD Anderson Cancer Center, Houston, USA. She received her Ph.D. in Bioinformatics from the University of North Carolina at Chapel Hill, USA. Her research focuses on characterizing the genetic and clinical diversity of patient tumors, and identifying and validating therapeutic targets and predictive molecular signatures through computational approaches.

Gordon B. Mills: At the University of Texas MD Anderson Cancer Center (UTMDACC), Houston, USA, Dr. Mills is the Chair of Systems Biology, Co-Director of the Institute for Personalized Cancer Therapy (IPCT), the Head of the Kleberg Center for Molecular Markers and Co-Director of the Ovarian Cancer Moonshot. The Kleberg Center for Molecular Markers is responsible for developing the markers needed for personalized molecular medicine and the IPCT is responsible for the implementation into clinical practice. The Ovarian Cancer Moonshot is tasked with developing transformational approaches to improving patient outcomes. Dr. Mills' research ranges across: mechanistic studies determining the role of genomic and other aberrations present in patient tumors; identification and validation of therapeutic targets; developing, validating, and implementing molecular markers; and integrating data through a cancer systems biology approach into robust predictive mathematical models.

Nidhi Sahni: Nidhi Sahni is an Assistant Professor of Systems Biology at the University of Texas MD Anderson Cancer Center (UTMDACC), Houston, USA. She is an affiliated faculty member of the Program in Structural and Computational Biology at Baylor College of Medicine, Houston, USA, and is a regular faculty member in the University of Texas Graduate School of Biomedical Sciences at Houston. She is also a CPRIT (Cancer Prevention Research Institute of Texas) scholar. She received her Ph.D. from the University of Iowa, Iowa City, USA, and did her postdoctoral work at Dana-Farber Cancer Institute, Boston, Massachusetts, USA. Her laboratory focuses on the systems biology of human cancer, integrating large-scale computational genomics and high-throughput experimental platforms to address fundamental problems in the modern era of personalized or precision medicine. The lab seeks a systems-level understanding of the underlying genetic and epigenetic aberrations in cancer heterogeneity and immunity.

Homepage: http://faculty.mdanderson.org/Nidhi_Sahni/Default.asp

Glossary terms

Missense mutations

(Also known as non-synonymous mutations). Nucleotide mutations in exons of protein-coding genes that cause amino-acid substitutions in the protein.

Frame-shift mutations

Nucleotide mutations in exons of protein-coding genes that cause an alteration to the reading frame of translation and usually result in a premature stop codon and a truncated or non-expressed protein. They typically involve small insertions or deletions of a number of nucleotides that is not divisible by 3.

Silent mutations

(Also known as synonymous mutations). Nucleotide mutations exons of protein-coding genes that do not alter the coded amino acid (due to degeneracy in the genetic code).

Nonsense mutations

Nucleotide mutations in exons of protein-coding genes that change amino-acid-encoding codons into stop codons.

ChIP-seq

(Chromatin immunoprecipitation followed by sequencing). Antibody-based immunoprecipitation of a chromatin-associated protein, such as a transcription factor (often epitope-tagged), and its potentially interacting crosslinked DNA fragments, followed by sequencing to reveal the identity of these DNA fragments. Overall, this approach reveals the genomic sites of occupancy of the protein of interest.

DNase-seq

Genome-wide sequencing of open chromatin regions that are sensitive to cleavage by DNase I. Open chromatin is enriched for regulatory sequences.

Chromosome conformation capture

A method that analyses the spatial organization of chromatin in a cell by quantifying the interactions between genomic loci that are in proximity in 3D space.

Gene ontology

A unified representation of attributes for genes and gene products across species, which helps functional interpretation of experimental data.

Topological centrality

In molecular interaction networks, topological centrality is an intrinsic network property that measures the overall position and connectedness of a node in the networks.

Nicking

Creating a single-strand DNA break.

References

1. Vogelstein B. Cancer genome landscapes *Science*. 2013; 339:1546–1558. [PubMed: 23539594]
2. Karchin R, Nussinov R. Genome Landscapes of Disease: Strategies to Predict the Phenotypic Consequences of Human Germline and Somatic Variation *PLoS Comput. Biol.* 2016; 12:e1005043. [PubMed: 27536867]
3. Vidal M, Cusick ME, Barabási AL. Interactome networks and human disease *Cell*. 2011; 144:986–998. [PubMed: 21414488]
4. Sahni N. Edgotype: a fundamental link between genotype and phenotype *Current opinion in genetics & development*. 2013; 23:649–657. [PubMed: 24287335]
5. Weinberg RA. Coming full circle—from endless complexity to simplicity and back again *Cell*. 2014; 157:267–271. [PubMed: 24679541]
6. Barabasi AL, Gulbahce N, Loscalzo J. *Nature reviews. Genetics*. 2011; 12:56–68.
7. Takiar V, Ip CK, Gao M, Mills GB, Cheung LW. Neomorphic mutations create therapeutic challenges in cancer *Oncogene*. 2016
8. Kim E. Systematic Functional Interrogation of Rare Cancer Variants Identifies Oncogenic Alleles *Cancer Discov.* 2016; 6:714–726. [PubMed: 27147599]
9. Rolland T. A proteome-scale map of the human interactome network *Cell*. 2014; 159:1212–1226. [PubMed: 25416956]
10. Huttlin EL. The BioPlex Network: A Systematic Exploration of the Human Interactome *Cell*. 2015; 162:425–440. [PubMed: 26186194]
11. Fuxman Bass JJ. Human gene-centered transcription factor networks for enhancers and disease variants *Cell*. 2015; 161:661–673. [PubMed: 25910213]
12. Sahni N. Widespread macromolecular interaction perturbations in human genetic disorders *Cell*. 2015; 161:647–660. [PubMed: 25910212]
13. Hofree M, Shen JP, Carter H, Gross A, Ideker T. Network-based stratification of tumor mutations *Nat. Methods*. 2013; 10:1108–1115. [PubMed: 24037242]
14. Ryan CJ. High-resolution network biology: connecting sequence with function *Nat. Rev. Genet.* 2013; 14:865–879. [PubMed: 24197012]
15. Boone C, Bussey H, Andrews BJ. Exploring genetic interactions and networks with yeast *Nat. Rev. Genet.* 2007; 8:437–449. [PubMed: 17510664]
16. Choudhary C, Mann M. Decoding signalling networks by mass spectrometry-based proteomics *Nat. Rev. Mol. Cell Biol.* 2010; 11:427–439. [PubMed: 20461098]
17. Khurana E. Role of non-coding sequence variants in cancer *Nature reviews. Genetics*. 2016; 17:93–108.
18. Futreal PA. A census of human cancer genes *Nat. Rev. Cancer*. 2004; 4:177–183. [PubMed: 14993899]
19. Baca SC. Punctuated evolution of prostate cancer genomes *Cell*. 2013; 153:666–677. [PubMed: 23622249]
20. Stephens PJ. Massive genomic rearrangement acquired in a single catastrophic event during cancer development *Cell*. 2011; 144:27–40. [PubMed: 21215367]
21. Nik-Zainal S. Mutational processes molding the genomes of 21 breast cancers *Cell*. 2012; 149:979–993. [PubMed: 22608084]
22. Knudson AG Jr. Mutation and cancer: statistical study of retinoblastoma *Proc Natl Acad Sci U S A*. 1971; 68:820–823. [PubMed: 5279523]
23. Nowell PC. The clonal evolution of tumor cell populations *Science*. 1976; 194:23–28. [PubMed: 959840]
24. Cairns J. Mutation selection and the natural history of cancer *Nature*. 1975; 255:197–200. [PubMed: 1143315]
25. Greaves M, Maley CC. Clonal evolution in cancer *Nature*. 2012; 481:306–313. [PubMed: 22258609]
26. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology *Trends Genet.* 2014; 30:418–426. [PubMed: 25108476]

27. Kandoth C. Mutational landscape and significance across 12 major cancer types *Nature*. 2013; 502:333–339. [PubMed: 24132290]
28. Kim NW. Specific association of human telomerase activity with immortal cells and cancer *Science*. 1994; 266:2011–2015. [PubMed: 7605428]
29. Huang FW. Highly recurrent TERT promoter mutations in human melanoma *Science*. 2013; 339:957–959. [PubMed: 23348506]
30. Horn S. TERT promoter mutations in familial and sporadic melanoma *Science*. 2013; 339:959–961. [PubMed: 23348503]
31. Borah S. Cancer. TERT promoter mutations and telomerase reactivation in urothelial cancer *Science (New York, NY)*. 2015; 347:1006–1010.
32. Bell RJ. Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer *Science*. 2015; 348:1036–1039. [PubMed: 25977370]
33. Buechner J, Einvik C. N-myc and noncoding RNAs in neuroblastoma *Mol. Cancer Res*. 2012; 10:1243–1253. [PubMed: 22936790]
34. Liu PY. Effects of a novel long noncoding RNA, lncUSMycN, on N-Myc expression and neuroblastoma progression *J. Natl. Cancer Inst*. 2014:106.
35. Calin GA. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia *Proc. Natl. Acad. Sci. U. S. A.* 2002; 99:15524–15529. [PubMed: 12434020]
36. Cancer Genome Atlas Research, N The Cancer Genome Atlas Pan-Cancer analysis project *Nat. Genet*. 2013; 45:1113–1120. [PubMed: 24071849]
37. Stratton MR, Campbell PJ, Futreal PA. The cancer genome *Nature*. 2009; 458:719–724. [PubMed: 19360079]
38. Garraway LA, Lander ES. Lessons from the cancer genome *Cell*. 2013; 153:17–37. [PubMed: 23540688]
39. Pakneshan S, Salajegheh A, Smith RA, Lam AK. Clinicopathological relevance of BRAF mutations in human cancer *Pathology*. 2013; 45:346–356. [PubMed: 23594689]
40. Ciriello G. Emerging landscape of oncogenic signatures across human cancers *Nat. Genet*. 2013; 45:1127–1133. [PubMed: 24071851]
41. Fujimoto A. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer *Nature genetics*. 2016; 48:500–509. [PubMed: 27064257]
42. Evans P, Avey S, Kong Y, Krauthammer M. Adjusting for background mutation frequency biases improves the identification of cancer driver genes *IEEE Trans Nanobioscience*. 2013; 12:150–157. [PubMed: 23694700]
43. Hodis E. A landscape of driver mutations in melanoma *Cell*. 2012; 150:251–263. [PubMed: 22817889]
44. Dees ND. MuSiC: identifying mutational significance in cancer genomes *Genome Res*. 2012; 22:1589–1598. [PubMed: 22759861]
45. Lawrence MS. Mutational heterogeneity in cancer and the search for new cancer-associated genes *Nature*. 2013; 499:214–218. [PubMed: 23770567]
46. Makova KD, Hardison RC. The effects of chromatin organization on variation in mutation rates in the genome *Nat. Rev. Genet*. 2015; 16:213–223. [PubMed: 25732611]
47. Sjoblom T. The consensus coding sequences of human breast and colorectal cancers *Science*. 2006; 314:268–274. [PubMed: 16959974]
48. Jones S. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses *Science*. 2008; 321:1801–1806. [PubMed: 18772397]
49. Parsons DW. An integrated genomic analysis of human glioblastoma multiforme *Science*. 2008; 321:1807–1812. [PubMed: 18772396]
50. Wood LD. The genomic landscapes of human breast and colorectal cancers *Science*. 2007; 318:1108–1113. [PubMed: 17932254]
51. Forbes SA. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer *Nucleic Acids Res*. 2011; 39:D945–950. [PubMed: 20952405]
52. Alberts, B.; *Molecular Biology of the Cell*. 4th. Garland Science; 2002.

53. MacArthur DG. Guidelines for investigating causality of sequence variants in human disease *Nature*. 2014; 508:469–476. [PubMed: 24759409]
54. Chakravarti A, Clark AG, Mootha VK. Distilling pathophysiology from complex disease genetics *Cell*. 2013; 155:21–26. [PubMed: 24074858]
55. Das J. Exploring mechanisms of human disease through structurally resolved protein interactome networks *Molecular bioSystems*. 2014; 10:9–17. [PubMed: 24096645]
56. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function *Nucleic Acids Res*. 2003; 31:3812–3814. [PubMed: 12824425]
57. Sim NL. SIFT web server: predicting effects of amino acid substitutions on proteins *Nucleic Acids Res*. 2012; 40:W452–457. [PubMed: 22689647]
58. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes *Nat. Protoc*. 2016; 11:1–9. [PubMed: 26633127]
59. Miller ML. Pan-Cancer Analysis of Mutation Hotspots in Protein Domains *Cell systems*. 2015; 1:197–209. [PubMed: 27135912]
60. Chen T. Hotspot mutations delineating diverse mutational signatures and biological utilities across cancer types *BMC Genomics*. 2016; 17(Suppl 2):394. [PubMed: 27356755]
61. Jia P. MSEA: detection and quantification of mutation hotspots through mutation set enrichment analysis *Genome Biol*. 2014; 15:489. [PubMed: 25348067]
62. Jones S. Personalized genomic analyses for cancer mutation discovery and interpretation *Sci. Transl. Med*. 2015; 7:283ra253.
63. Gudbjartsson DF. Large-scale whole-genome sequencing of the Icelandic population *Nat. Genet*. 2015; 47:435–444. [PubMed: 25807286]
64. Adzhubei IA. A method and server for predicting damaging missense mutations *Nat. Methods*. 2010; 7:248–249. [PubMed: 20354512]
65. Quan L, Lv Q, Zhang Y. STRUM: structure-based prediction of protein stability changes upon single-point mutation *Bioinformatics*. 2016; 32:2936–2946. [PubMed: 27318206]
66. Porta-Pardo E, Godzik A. e-Driver: a novel method to identify protein regions driving cancer *Bioinformatics*. 2014; 30:3109–3114. [PubMed: 25064568]
67. Gonzalez-Perez A. Computational approaches to identify functional genetic variants in cancer genomes *Nature*. 2013; 10:723–729.
68. Ding L, Wendl MC, McMichael JF, Raphael BJ. Expanding the computational toolbox for mining cancer genomes *Nat. Rev. Genet*. 2014; 15:556–570. [PubMed: 25001846]
69. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics *Nucleic Acids Res*. 2011; 39:e118. [PubMed: 21727090]
70. Reimand J, Bader GD. Systematic analysis of somatic mutations in phosphorylation signaling predicts novel cancer drivers *Mol. Syst. Biol*. 2013; 9:637. [PubMed: 23340843]
71. Creixell P. Kinome-wide decoding of network-attacking mutations rewiring cancer signaling *Cell*. 2015; 163:202–217. [PubMed: 26388441]
72. Creixell P. Unmasking determinants of specificity in the human kinome *Cell*. 2015; 163:187–201. [PubMed: 26388442]
73. Meyer MJ. mutation3D: Cancer Gene Prediction Through Atomic Clustering of Coding Variants in the Structural Proteome *Hum. Mutat*. 2016; 37:447–456. [PubMed: 26841357]
74. Niu B. Protein-structure-guided discovery of functional mutations across 19 cancer types *Nature genetics*. 2016; 48:827–837. [PubMed: 27294619]
75. Wang X. Three-dimensional reconstruction of protein networks provides insight into human genetic disease *Nature biotechnology*. 2012; 30:159–164.
76. Mao Y. CanDrA: cancer-specific driver missense mutation annotation with optimized features *PLoS One*. 2013; 8:e77945. [PubMed: 24205039]
77. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations *Nat. Methods*. 2010; 7:575–576. [PubMed: 20676075]
78. Ghersi D, Singh M. Interaction-based discovery of functionally important genes in cancers *Nucleic Acids Res*. 2014; 42:e18. [PubMed: 24362839]

79. Zhao J, Cheng F, Wang Y, Arteaga CL, Zhao Z. Systematic Prioritization of Druggable Mutations in approximately 5000 Genomes Across 16 Cancer Types Using a Structural Genomics-based Approach *Mol. Cell. Proteomics*. 2016; 15:642–656. [PubMed: 26657081]
80. An integrated encyclopedia of DNA elements in the human genome *Nature*. 2012; 489:57–74. [PubMed: 22955616]
81. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR *Nat. Protoc*. 2015; 10:1556–1566. [PubMed: 26379229]
82. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data *Nucleic Acids Res*. 2010; 38:e164. [PubMed: 20601685]
83. Kircher M. A general framework for estimating the relative pathogenicity of human genetic variants *Nature genetics*. 2014; 46:310–315. [PubMed: 24487276]
84. Ritchie GR, Dunham I, Zeggini E, Flicek P. Functional annotation of noncoding sequence variants *Nat. Methods*. 2014; 11:294–296. [PubMed: 24487584]
85. Gulko B, Hubisz MJ, Gronau I, Siepel A. A method for calculating probabilities of fitness consequences for point mutations across the human genome *Nat. Genet*. 2015; 47:276–283. [PubMed: 25599402]
86. Lee D. A method to predict the impact of regulatory variants from DNA sequence *Nat. Genet*. 2015; 47:955–961. [PubMed: 26075791]
87. Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning-based sequence model *Nat. Methods*. 2015; 12:931–934. [PubMed: 26301843]
88. Li MJ, Wang LY, Xia Z, Sham PC, Wang J. GWAS3D: Detecting human regulatory variants by integrative analysis of genome-wide associations, chromosome interactions and histone modifications *Nucleic Acids Res*. 2013; 41:W150–158. [PubMed: 23723249]
89. Gonzalez-Perez A, Deu-Pons J, Lopez-Bigas N. Improving the prediction of the functional impact of cancer mutations by baseline tolerance transformation *Genome Med*. 2012; 4:89. [PubMed: 23181723]
90. Kaminker JS, Zhang Y, Watanabe C, Zhang Z. CanPredict: a computational tool for predicting cancer-associated missense mutations *Nucleic Acids Res*. 2007; 35:W595–598. [PubMed: 17537827]
91. Yates CM, Filippis I, Kelley LA, Sternberg MJ. SuSPect: enhanced prediction of single amino acid variant (SAV) phenotype using network features *J. Mol. Biol*. 2014; 426:2692–2701. [PubMed: 24810707]
92. Fu Y. FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer *Genome Biol*. 2014; 15:480. [PubMed: 25273974]
93. Chen JC. Identification of causal genetic drivers of human disease through systems-level analysis of regulatory networks *Cell*. 2014; 159:402–414. [PubMed: 25303533]
94. Bashashati A. DriverNet: uncovering the impact of somatic driver mutations on transcriptional networks in cancer *Genome Biol*. 2012; 13:R124. [PubMed: 23383675]
95. Paull EO. Discovering causal pathways linking genomic events to transcriptional states using Tied Diffusion Through Interacting Events (TieDIE) *Bioinformatics*. 2013; 29:2757–2764. [PubMed: 23986566]
96. Bertrand D. Patient-specific driver gene prediction and risk assessment through integrated network analysis of cancer omics profiles *Nucleic Acids Res*. 2015; 43:e44. [PubMed: 25572314]
97. Jia P, Zhao Z. VarWalker: personalized mutation network analysis of putative cancer genes from next-generation sequencing data *PLoS computational biology*. 2014; 10:e1003460. [PubMed: 24516372]
98. Vandin F, Upfal E, Raphael BJ. Algorithms for detecting significantly mutated pathways in cancer *J. Comput. Biol*. 2011; 18:507–522. [PubMed: 21385051]
99. Leiserson MD. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes *Nat. Genet*. 2015; 47:106–114. [PubMed: 25501392]
100. Dehouck Y, Kwasigroch JM, Rومان M, Gilis D. BeAtMuSiC: Prediction of changes in protein-protein binding affinity on mutations *Nucleic Acids Res*. 2013; 41:W333–339. [PubMed: 23723246]

101. Vazquez M, Valencia A, Pons T. Structure-PPI: a module for the annotation of cancer-related single-nucleotide variants at protein-protein interfaces *Bioinformatics*. 2015; 31:2397–2399. [PubMed: 25765346]
102. Li M, Simonetti FL, Goncarenco A, Panchenko AR. MutaBind estimates and interprets the effects of sequence variants on protein-protein interactions *Nucleic Acids Res*. 2016; 44:W494–501. [PubMed: 27150810]
103. Mosca R. dSysMap: exploring the edgetic role of disease mutations *Nat. Methods*. 2015; 12:167–168. [PubMed: 25719824]
104. Mosca R, Ceol A, Aloy P. Interactome3D: adding structural details to protein networks *Nat. Methods*. 2013; 10:47–53. [PubMed: 23399932]
105. Di Minin G. Mutant p53 reprograms TNF signaling in cancer cells through interaction with the tumor suppressor DAB2IP *Mol. Cell*. 2014; 56:617–629. [PubMed: 25454946]
106. Reece-Hoyes JS. Extensive rewiring and complex evolutionary dynamics in a *C. elegans* multiparameter transcription factor network *Mol. Cell*. 2013; 51:116–127. [PubMed: 23791784]
107. Barrera LA. Survey of variation in human transcription factors reveals prevalent DNA binding changes *Science*. 2016; 351:1450–1454. [PubMed: 27013732]
108. Macintyre G, Bailey J, Haviv I, Kowalczyk A. is-rSNP: a novel technique for in silico regulatory SNP detection *Bioinformatics*. 2010; 26:i524–530. [PubMed: 20823317]
109. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants *Nucleic Acids Res*. 2012; 40:D930–934. [PubMed: 22064851]
110. Perera D. OncoCis: annotation of cis-regulatory mutations in cancer *Genome Biol*. 2014; 15:485. [PubMed: 25298093]
111. Wang J, Batmanov K. BayesPI-BAR: a new biophysical model for characterization of regulatory sequence variations *Nucleic Acids Res*. 2015; 43:e147. [PubMed: 26202972]
112. Jonas S, Izaurralde E. Towards a molecular understanding of microRNA-mediated gene silencing *Nat. Rev. Genet*. 2015; 16:421–433. [PubMed: 26077373]
113. Krol J, Loedige I, Filipowicz W. The widespread regulation of microRNA biogenesis, function and decay *Nat. Rev. Genet*. 2010; 11:597–610. [PubMed: 20661255]
114. Hiard S, Charlier C, Coppieters W, Georges M, Baurain D. Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates *Nucleic Acids Res*. 2010; 38:D640–651. [PubMed: 19906729]
115. Bhattacharya A, Ziebarth JD, Cui Y. SomamiR: a database for somatic mutations impacting microRNA function in cancer *Nucleic Acids Res*. 2013; 41:D977–982. [PubMed: 23180788]
116. Bao L. PolymiRTS Database: linking polymorphisms in microRNA target sites with complex traits *Nucleic Acids Res*. 2007; 35:D51–54. [PubMed: 17099235]
117. Ziebarth JD, Bhattacharya A, Chen A, Cui Y. PolymiRTS Database 2.0: linking polymorphisms in microRNA target sites with human diseases and complex traits *Nucleic Acids Res*. 2012; 40:D216–221. [PubMed: 22080514]
118. Bhattacharya A, Ziebarth JD, Cui Y. PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways *Nucleic Acids Res*. 2014; 42:D86–91. [PubMed: 24163105]
119. Shalek AK. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation *Nature*. 2014; 510:363–369. [PubMed: 24919153]
120. Duan Q. LINCS Canvas Browser: interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures *Nucleic Acids Res*. 2014; 42:W449–460. [PubMed: 24906883]
121. Dogruluk T. Identification of Variant-Specific Functions of PIK3CA by Rapid Phenotyping of Rare Mutations *Cancer Res*. 2015; 75:5341–5354. [PubMed: 26627007]
122. Cheung LW. Regulation of the PI3K pathway through a p85alpha monomer-homodimer equilibrium *Elife*. 2015; 4:e06866. [PubMed: 26222500]
123. Cheung LW. Naturally occurring neomorphic PIK3R1 mutations activate the MAPK pathway, dictating therapeutic response to MAPK pathway inhibitors *Cancer Cell*. 2014; 26:479–494. [PubMed: 25284480]

124. Cheung LW. High frequency of PIK3R1 and PIK3R2 mutations in endometrial cancer elucidates a novel mechanism for regulation of PTEN protein stability *Cancer Discov.* 2011; 1:170–185. [PubMed: 21984976]
125. Lambert JP. Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition *Nat. Methods.* 2013; 10:1239–1245. [PubMed: 24162924]
126. Mertins P. Proteogenomics connects somatic mutations to signalling in breast cancer *Nature.* 2016; 534:55–62. [PubMed: 27251275]
127. Zhang H. Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer *Cell.* 2016; 166:755–765. [PubMed: 27372738]
128. Macias E. An ARF-independent c-MYC-activated tumor suppression pathway mediated by ribosomal protein-Mdm2 Interaction *Cancer Cell.* 2010; 18:231–243. [PubMed: 20832751]
129. Zhong Q. Edgetic perturbation models of human inherited disorders *Mol. Syst. Biol.* 2009; 5:321. [PubMed: 19888216]
130. Fields S, Song O. A novel genetic system to detect protein-protein interactions *Nature.* 1989; 340:245–246. [PubMed: 2547163]
131. Cassonnet P. Benchmarking a luciferase complementation assay for detecting protein complexes *Nat. Methods.* 2011; 8:990–992. [PubMed: 22127214]
132. Osato M. Point mutations in the RUNX1/AML1 gene: another actor in RUNX leukemia *Oncogene.* 2004; 23:4284–4296. [PubMed: 15156185]
133. Gerstein MB. Architecture of the human regulatory network derived from ENCODE data *Nature.* 2012; 489:91–100. [PubMed: 22955619]
134. Landt SG. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia *Genome Res.* 2012; 22:1813–1831. [PubMed: 22955991]
135. Hu S. Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling *Cell.* 2009; 139:610–622. [PubMed: 19879846]
136. Weirauch MT. Determination and inference of eukaryotic transcription factor sequence specificity *Cell.* 2014; 158:1431–1443. [PubMed: 25215497]
137. Deplancke B, Dupuy D, Vidal M, Walhout AJ. A gateway-compatible yeast one-hybrid system *Genome Res.* 2004; 14:2093–2101. [PubMed: 15489331]
138. Ding J. Systematic analysis of somatic mutations impacting gene expression in 12 tumour types *Nat Commun.* 2015; 6:8554. [PubMed: 26436532]
139. Ng S. PARADIGM-SHIFT predicts the function of mutations in multiple cancers using pathway impact analysis *Bioinformatics.* 2012; 28:i640–i646. [PubMed: 22962493]
140. Taguchi A. Proteomic signatures associated with p53 mutational status in lung adenocarcinoma *Proteomics.* 2014; 14:2750–2759. [PubMed: 25331784]
141. Cong L. Multiplex genome engineering using CRISPR/Cas systems *Science (New York, NY).* 2013; 339:819–823.
142. Mali P. RNA-guided human genome engineering via Cas9 *Science (New York, NY).* 2013; 339:823–826.
143. Jinek M. RNA-programmed genome editing in human cells *eLife.* 2013; 2:e00471. [PubMed: 23386978]
144. Cho SW, Kim S, Kim JM, Kim J-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease *Nat. Biotechnol.* 2013; 31:230–232. [PubMed: 23360966]
145. Jiang W, Marraffini LA. CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial Immunity Systems *Annu. Rev. Microbiol.* 2015; 69:209–228. [PubMed: 26209264]
146. Hsu PD. DNA targeting specificity of RNA-guided Cas9 nucleases *Nat Biotechnol.* 2013; 31:827–832. [PubMed: 23873081]
147. Shalem O, Sanjana NE, Zhang F. High-throughput functional genomics using CRISPR-Cas9 *Nat. Rev. Genet.* 2015; 16:299–311. [PubMed: 25854182]
148. Shalem O. Genome-scale CRISPR-Cas9 knockout screening in human cells *Science.* 2014; 343:84–87. [PubMed: 24336571]

149. Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic screens in human cells using the CRISPR-Cas9 system *Science*. 2014; 343:80–84. [PubMed: 24336569]
150. Koike-Yusa H, Li Y, Tan EP, Velasco-Herrera Mdel C, Yusa K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library *Nat. Biotechnol.* 2014; 32:267–273. [PubMed: 24535568]
151. Wang T. Identification and characterization of essential genes in the human genome *Science*. 2015; 350:1096–1101. [PubMed: 26472758]
152. Doench JG. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9 *Nat. Biotechnol.* 2016; 34:184–191. [PubMed: 26780180]
153. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening *Nat. Methods*. 2014; 11:783–784. [PubMed: 25075903]
154. Ma H. A CRISPR-Based Screen Identifies Genes Essential for West-Nile-Virus-Induced Cell Death *Cell Rep.* 2015; 12:673–683. [PubMed: 26190106]
155. Hart T. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities *Cell*. 2015; 163:1515–1526. [PubMed: 26627737]
156. Maruyama T. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining *Nat Biotechnol.* 2015; 33:538–542. [PubMed: 25798939]
157. Chu VT. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells *Nat Biotechnol.* 2015; 33:543–548. [PubMed: 25803306]
158. Paquet D. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9 *Nature*. 2016; 533:125–129. [PubMed: 27120160]
159. Garst AD. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering *Nat. Biotechnol.* 2017; 35:48–55. [PubMed: 27941803]
160. Gutschner T, Haemmerle M, Genovese G, Draetta GF, Chin L. Post-translational Regulation of Cas9 during G1 Enhances Homology-Directed Repair *Cell Rep.* 2016; 14:1555–1566. [PubMed: 26854237]
161. Yu C. Small molecules enhance CRISPR genome editing in pluripotent stem cells *Cell stem cell*. 2015; 16:142–147. [PubMed: 25658371]
162. Ran FA. In vivo genome editing using *Staphylococcus aureus* Cas9 *Nature*. 2015; 520:186–191. [PubMed: 25830891]
163. Yang Y. A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice *Nat Biotechnol.* 2016; 34:334–338. [PubMed: 26829317]
164. Nishida K. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems *Science*. 2016:353.
165. Hess GT. Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells *Nat Methods*. 2016; 13:1036–1042. [PubMed: 27798611]
166. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage *Nature*. 2016; 533:420–424. [PubMed: 27096365]

Reference highlights

6. Barabasi, A.L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).

Shows network models for molecular and pathway relationships for complex diseases.

7. Takiar, V., Ip, C.K., Gao, M., Mills, G.B. & Cheung, L.W. Neomorphic mutations create therapeutic challenges in cancer. *Oncogene* (2016).

Highlights diverse functional effects of different edgetic or neomorphic mutations, which should be taken into account for designing precision medicine.

8. Kim, E. et al. Systematic Functional Interrogation of Rare Cancer Variants Identifies Oncogenic Alleles. *Cancer Discov.* **6**, 714–726 (2016).

One of the first papers showing systematic characterization of distinct cancer hallmark behaviors of rare oncogenic alleles.

9. Rolland, T. et al. A proteome-scale map of the human interactome network. *Cell* **159**, 1212–1226 (2014).

One of the largest scale human interactome network maps identifying novel connectivity modules between cancer proteins.

11. Fuxman Bass, J.I. et al. Human gene-centered transcription factor networks for enhancers and disease variants. *Cell* **161**, 661–673 (2015).

One of the first studies to characterize the protein–DNA interaction altered by enhancer mutations at large scale.

12. Sahni, N. et al. Widespread macromolecular interaction perturbations in human genetic disorders. *Cell* **161**, 647–660 (2015).

One of the first papers showing systematic characterization of a large number of mutations involved in ~1,000 human diseases, in terms of their functional effect on protein–protein and protein–DNA interaction networks, and protein folding/stability.

13. Hofree, M., Shen, J.P., Carter, H., Gross, A. & Ideker, T. Network-based stratification of tumor mutations. *Nat. Methods* **10**, 1108–1115 (2013).

A method that integrates tumor genomes with gene networks to cluster together patients with mutations in similar network neighborhood.

17. Khurana, E. et al. Role of non-coding sequence variants in cancer. *Nat. Rev. Genet.* **17**, 93–108 (2016).

Shows recent computational and experimental advances in evaluating the functional impact of non-coding cancer variants.

103. Mosca, R. et al. dSysMap: exploring the edgetic role of disease mutations. *Nat. Methods* **12**, 167–168 (2015).

Method that maps missense disease mutations onto the structurally resolved human interactome.

125. Lambert, J.P. et al. Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition. *Nat. Methods* **10**, 1239–1245 (2013).

This paper assesses the alterations in protein interaction partners by AP-MS.

Key Points

- Current cancer therapies targeted to particular genetic lesions are primarily hampered by the extreme genetic heterogeneity observed across patient populations.
- Cancer genomic variants and regulatory molecules interact with each other in cellular networks. Network biology has recently emerged as a systems-level approach to stratifying mutations that give rise to markedly different phenotypes.
- Different cancer mutations often lead to distinct perturbations in signal transduction networks.
- A number of computational tools have been developed to analyze functional impacts of cancer mutations and to prioritize drivers of oncogenesis.
- Various experimental strategies have emerged to study mutation-specific functional effects. Such functional variomics approaches can dissect cancer variants at high resolution.
- Integration of computational predictions with systems biology experimental approaches will be critical for interpreting complex genotype-to-phenotype relationships in human disease including cancer. Together this effort represents a critical step towards precision medicine.

Figure 1 |. Complex genetic heterogeneity in human cancer

a | Rapid increase in the number of cancer mutations identified over the past decade.

Mutations were downloaded from the COSMIC database and Pubmed IDs for each mutation were extracted. The publication year for each Pubmed ID was obtained by 'RISmed' R package. The number of mutations was plotted as a function of the corresponding publication year. **b** | Mutations occur in both coding and non-coding regions of cancer genomes. Coding mutations are located in genes that undergo mRNA transcription and protein translation. Non-coding aberrations include mutations in *cis*-regulatory elements and in non-coding RNAs. Furthermore, mutations can be small-scale point mutations up to larger-scale aberrations such as copy-number variation and chromosomal rearrangements (not shown). RNAPII, RNA polymerase II. **c** | Heterogeneous coding mutations across

diverse cancer types. Somatic mutations across 33 types of cancer were obtained from The Cancer Genome Atlas (TCGA) project, comprised of 10,489 tumor samples. The number of mutations per sample was plotted as boxplot, with the x-axis corresponding to cancer types and the y-axis representing the log10 number of mutations per sample. Cancer types are ordered from left to right based on tissue origin.

ACC, adrenocortical carcinoma; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, cholangiocarcinoma; COAD, colon adenocarcinoma; DLBC, lymphoid neoplasm diffuse large b-cell lymphoma; ESCA, esophageal carcinoma; GBM, glioblastoma multiforme; HNSC, head and neck squamous carcinoma; KICH, kidney chromophobe; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LAML, acute myeloid leukemia; LGG, brain lower grade glioma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; MESO, mesothelioma; OV, ovarian serous cystadenocarcinoma; PAAD, pancreatic adenocarcinoma; PCPG, pheochromocytoma and paraganglioma; PRAD, prostate adenocarcinoma; READ, rectum adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; TGCT, testicular germ cell tumors; THCA, thyroid carcinoma; THYM, thymoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma; UVM, uveal melanoma.

Figure 3 | Effects of cancer variants on molecular interaction networks in cells

a | Schematic illustration of distinct molecular interaction profiles caused by heterogeneous genetic mutations. Nodes are macromolecules such as proteins, DNAs and RNAs, whereas edges are biophysical interactions between them. Solid lines represent retained interactions and dashed lines represent perturbed interactions by mutations. Cancer-associated mutations can cause a wide range of effects on cellular interaction networks, including loss of all interactions, edgetic perturbation of specific interaction(s), and edgetic gain of interaction(s).

b | Residues affected by mutations are highlighted on the CDK4 structure based on homology modelling (Protein data bank (PDB) ID=1bi7). Mutation locations on the protein structure are also shown.

c | Residues affected by mutations are highlighted on the FBP1 structure (PDB ID= 1fpi). Mutation locations on the protein structure are also shown.

Figure 4 |. Computational tools that prioritize cancer genes and mutations

Coding genes and non-coding regulatory elements interact with each other in cellular networks (centre). Cancer-associated mutations from The Cancer Genome Atlas (TCGA), the International Cancer Genome Consortium (ICGC) and other sources can be computationally prioritized as driver candidates, based on distinct features, including sequence (top left), structure (top right), regulatory features (middle left), functional and network features (middle right). Sequence features include conservation, mutation frequency, etc. Structural features include the 3D structural configuration of proteins, free energy changes, etc. Regulatory features include *cis*-regulatory sequences annotated by the Encyclopedia of DNA Elements (ENCODE) and its gene-annotation sub-project, GENCODE. Functional and network features include gene ontology, pathway and network centrality measurements. Cancer-causing driver mutations can also be prioritized by their effects on rewiring networks, including protein–protein, transcription factor (TF)–gene, miRNA–gene interaction networks. A list of computational tools is shown in the relevant white boxes.

Figure 5 |. Experimental platforms to characterize cancer mutations

Experimental pipelines for systematically characterizing functional changes induced by patient-specific mutations. Mutant clones are usually tested in parallel with their wild-type (WT) counterparts for comparative purposes. **a** | RNA sequencing (RNA-seq) compares the transcriptomic profiles of wild-type and mutant cells, based on next-generation sequencing of their respective cDNA pools. **b** | Library of integrated network-based cellular signatures (LINCS) L1000 compares the transcriptomic profiles of wild-type and mutant cells, based on a biotin and SAPE (streptavidin, phycoerythrin conjugated) optical detection system. **c** | Reverse-phase protein array (RPPA). RPPA is a high-throughput protein microarray technology that detects protein expression levels in tissue or cell lysates (wild-type or mutant) based on specific antibodies. **d** | Affinity purification coupled with mass spectrometry (AP-MS) identifies changes in protein interaction partners between wild-type and mutant proteins, based on antibody-based affinity purification, followed by mass spectrometry. The type of mass spectrometry indicated is liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) **e** | Enhanced yeast two-hybrid (eY2H). The DNA-binding domain (DB) and activation domain (AD) each fused to bait and prey proteins, respectively, reconstitute the transcription factor activity if brought in close proximity by the interaction between bait and prey proteins. **f** | Protein fragment complementation assay (PCA). Two complementary protein fragments fused to bait and prey proteins, respectively,

reconstitute the full-length functional protein when brought in close proximity by the interaction between bait and prey proteins. **g** | ChIP-seq combines chromatin immunoprecipitation (ChIP) with next-generation sequencing to identify DNA sites that proteins (such as transcription factors (TFs)) bind to. **h** | Enhanced yeast one-hybrid (eY1H). A DNA fragment is cloned upstream of a reporter. Upon binding of the protein of interest, the reporter is turned on and the activity can be measured. **i** | Protein-binding microarray (PBM). PBM is an *in vitro* technology that detects a broad spectrum of DNA-binding specificities for TFs at large scale, based on fluorescence measurement. **j** | CRISPR. The CRISPR system consists of two components: the CRISPR-associated endonuclease 9 (Cas9) and the single guide RNA (sgRNA). The specificity of the endonuclease is determined by the complementarity of the sgRNA and its 20-nucleotide target sequence in the genome. The Cas9 endonuclease creates DNA double-strand breaks (DSBs) at the target site, which are repaired either by the non-homologous end-joining (NHEJ) mechanism generating gene 'knockouts', or by homology-directed repair (HDR) for precise editing of the genome. Generated mutant strains can be tested downstream for alterations in molecular interactions, gene expression, fitness and drug resistance/sensitivity.

Table 1 |

Computational algorithms to prioritize cancer variants with functional impact

Algorithm	Type	Conservation	Structure	Coding	Network	Scoring	Annotation	Reference
<i>Sequence feature algorithms</i>								
SIFT/SIFT4G	SNV	1	0	1	0	1	1	58
MutSig/ MutSigCV	SNV, indel	1	0	1	0	1	0	45
MuSiC	SNV, indel	0	0	1	0	1	0	44
MSEA	SNV, indel	0	0	1	0	1	0	61
<i>Structure feature algorithms</i>								
PolyPhen-2	SNV	1	1	1	0	1	1	64
STRUM	SNV	1	1	1	0	1	0	65
CanDrA	SNV	1	1	1	0	1	1	76
MutationTaster	SNV	1	1	1	0	1	1	77
ActiveDriver	SNV	0	1	1	0	1	0	70
CanBind	SNV	0	1	1	0	1	0	78
SGDriver	SNV	0	1	1	0	1	0	79
e-Driver	SNV	0	1	1	0	1	0	66
Mutation3D	SNV	0	1	1	0	1	0	73
<i>Regulatory feature algorithms</i>								
ANNOVAR	SNV, indel, SV	1	1	1	0	0	1	82
CADD	SNV, indel	1	0	1	0	1	1	83
GWAVA	SNV, indel	1	0	1	0	1	1	84
FitCons	SNV	1	0	1	0	1	1	85
deltaSVM	SNV, indel	0	0	1	0	1	0	86
GWAS3D	SNV	1	0	1	0	1	1	88
DeepSEA	SNV, indel	1	0	1	0	1	1	87
<i>Gene ontology and network-based algorithms</i>								
CanPredict	SNV	0	0	1	1	1	0	90
transFIC	SNV	1	0	1	1	1	0	89
FunSeq2	SNV, indel	1	0	1	1	1	1	92
SuSPect	SNV	1	1	1	1	1	0	91
<i>Protein–protein interaction algorithms</i>								
BeAtMuSiC	SNV	0	1	1	0	1	0	100
Structure-PPi	SNV	1	1	1	0	1	0	101
dSysMap	SNV	0	1	1	0	0	0	103
MutaBind	SNV	0	1	1	0	1	0	102
<i>TF–gene interaction algorithms</i>								
is-rSNP	SNV	0	0	1	0	1	0	108
HaploReg	SNV, indel	1	0	1	0	0	1	109
OncoCis	SNV, indel	1	0	1	0	0	1	110

Algorithm	Type	Conservation	Structure	Coding	Network	Scoring	Annotation	Reference
BayesPI-BAR	SNV	1	1	1	0	1	0	111
<i>MicroRNA–gene interaction algorithms</i>								
Patrocles	SNV, SV	1	1	1	0	1	0	114
SomamiR	SNV, indel	0	0	1	0	0	0	115
PolymiRTS	SNV, indel	0	0	1	0	0	0	118
<i>Algorithms for identifying mutated subnetworks or pathways</i>								
DriverNet	SNV, indel, SV	0	0	1	1	1	0	94
TieDIE	SNV, indel, SV	0	0	1	1	1	0	95
OncoIMPACT	SNV, indel, SV	0	0	1	1	1	0	96
VarWalker	SNV, indel	0	0	1	1	1	0	97
HotNet/HotNet2	SNV, indel, SV	0	0	1	1	1	0	99
NBS	SNV, indel	0	0	1	1	1	0	13

SNV, single nucleotide variant; indel, small insertion or deletion; SV, structural variant; TF, transcription factor. ‘Coding’ indicates if the algorithm can analyze coding mutations; ‘Network’ indicates if the algorithm takes into account network properties; 1 or 0 represents if an algorithm includes or excludes a specific feature, respectively. For an extended version of this table, including descriptions of the tools, see Supplementary information S1 (table).