# 13 Systems Biology and Integrated Computational Methods for Cancer-Associated Mutation Analysis

Ayisha Zia and Sajid Rashid

## 13.1 Systems Biology

Systems biology is the research endeavor that offers the basic scientific groundwork for synthetic biology. It is grounded on the molecular diversity of living systems [1]. It is an integrative system that connects different components in a single biological unit and also links different units like cells and tissues using holistic methods to characterize their functions through computational methods, quantitative approaches, and high-throughput technologies. Cells are made of different constituents that interact and make a network model, for example, metabolic, regulatory, and signaling networks that regulate various cellular functions. Several elaborated and dynamic models are available for signaling pathways [2].

The computational approaches deliver a comprehension to understand the dynamics and interaction within cells, organs, tissues, and organisms. For complex diseases, precision medicine and quantitative methods are influenced by systems biology [3]. The best example of system thinking is the Human Genome Project as it shows different ways to work on the problems in the field of genetics [4]. Its main purpose is to discover the properties of cells, tissues, and organisms working as a whole system whose description is possible only by using systems biology which involves metabolic networks [5]. Interpretation of the systems biology to obtain and investigate complex data sets by interdisciplinary tools and experimental studies generally starts with omics including genomics, transcriptomics, and metabolomics. Other subdisciplines include phosphoproteomics, glycoproteomics, and areas to identify chemically modified proteins, metabolomics-, organismal-, tissue-, or cellular-level measurements of lipids [6].

A. Zia · S. Rashid (✉)
National Centre for Bioinformatics, Functional Informatics Lab, Quaid-i-Azam University, Islamabad, Pakistan
e-mail: azia@bs.qau.edu.pk; sajidrwp@yahoo.co.uk



N. Masood, S. Shakil Malik (eds.), *'Essentials of Cancer Genomic, Computational Approaches and Precision Medicine*,
https://doi.org/10.1007/978-981-15-1067-0_13

### 13.1.1 Dynamic State Characterization

For complex reaction networks, dynamic analysis involves tracing of time-dependent concentration changes and reaction fluxes over the time period. The three key points of dynamic states are:

- Time constant (the rate of change of a variable is considered by time constant)
- Aggregate variables (the biochemical and physiological events that are involved in unfolding. Basically, we move to the aggregate variables from the original concentration variables that eventually terminate in the overall dynamic features on slower time scales)
- Transitions (intricate networks can transition from one state to another state)

### 13.1.2 Formulation of Dynamic Network Models

Two different approaches (bottom-up and top-down) are used to formulate dynamic networks. In bottom-up approach, we identify all the events in the network in complexity with the addition of more information from time to time to make the event complete, whereas in top-down approach, all data and information is collected at the same time, and later this data is divided into smaller parts. The bottom-up analysis of any dynamic state of a network is based on the kinetic theory and network topology [2].

### 13.1.3 Cancer Systems Biology

It comprehends the application of systems biology methods to study the disease with evolving properties at different biological levels. It also helps to analyze how the disturbance in the intracellular pathways and networks of normal cells occurs during carcinogenesis for the development of effective prognostic models. These models can assist scientists in the validations of new treatments and drugs [7]. These perturbations are caused by the instability in tumors that changes the functions of different molecules. It is further convoluted due to the networks in a single cell and by the alterations in the interactions with the environment and whole individual during the tumorigenic process itself. Therefore computational and mathematical methods are used in cancer systems biology to interpret the complexity [8].

Cancer systems biology combines basic and clinical cancer research, and it provides applications of systems biology methods to the cancer research, particularly:

a. The need for improved methods to gain understanding from extensive networks
b. The significance of assimilating multiple types of data in construction of further accurate models
c. Trials in deciphering insights of tumorigenic mechanisms into therapeutic mediations
d. The function of tumor microenvironment at different levels [9]

### 13.1.4 Analysis of Cells at System Level

The system level provides the information at all the points of cell function. Different technologies are present which are supporting huge data quantity. This data needs to be handled and managed to make it into meaningful knowledge. Hence, organization and scrutinization of all data sets are known as bioinformatics like omics that delivers large amount of information from proteins, metabolites, and mRNA. By utilizing the sequencing technology, one can find genomic sequence and depict its determinants that include single nucleotide polymorphisms (SNPs) and regulatory sites that control particular phenotype or its function in an organism [10]. Epigenomics describes the epigenetic modifications [11], and proteomics measures the proteins and posttranslational modifications [12]. Likewise, transcriptomics measures transcriptome [11], and the study of metabolites in the cells and tissues is metabolomics [13]. These experimental techniques may result in huge data collection that may become a basis for designing novel tools and algorithms to interpret unknown data sets through knowledge-based information and linking the outcome of system-level studies.

## 13.2 Computational Approaches Used in Systems Biology

The data obtained from omics is organized by various tools of bioinformatics. These data sets are then used to form networks. Topological features can be constructed from these networks of molecular interactions. In these networks, there are network motifs [14] and functional modules that can perform functional tasks and represent dynamical signal properties. Regulatory pathways that include different motifs, feedback loops, and modules could be mined to construct dynamical models [15] which are further used for simulations to understand their promising behavior in time and space. To study the drug actions, they can be combined with PK/PD models [16]. Figure 13.1 shows the computational approaches used in systems biology.

### 13.2.1 Systems Medicine

Systems medicine is based on systems biology and systems science. It reflects intricate interactions within the human body with respect to the genome, environment, and behavior of the patient [17]. Systems medicine is used in research setups as it uncovers the unique and dynamic network of interactions which are crucial for influencing the progress of medical conditions. It also assists in determining molecular targets against any condition for its therapeutic and diagnostic measures. The relationship between the dry and wet lab is supported by systems medicine as well [18]. The basic dissimilarity among systems biology and systems medicine is that systems biology assumes the data to be useable and correct, while systems medicine ensures to lead with molecular and clinical data sets to produce the pathways that might contribute to medicinal development to the adapted healthcare [19, 20]. The

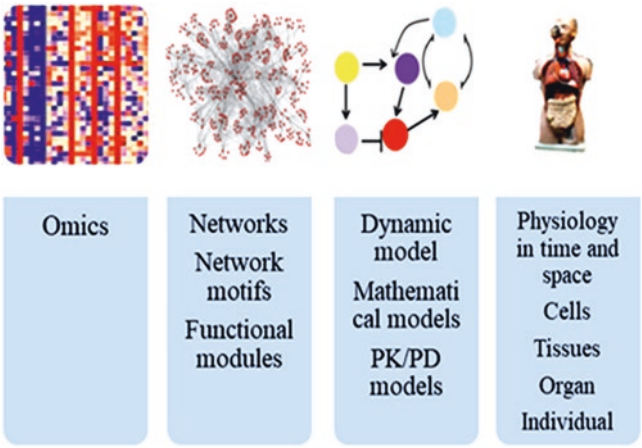


**Fig. 13.1** Computational approaches in systems biology. Omics data are arranged using different tools of bioinformatics to construct networks. Regulatory pathways can be extracted from these networks for the formation of dynamic models. From dynamic models, the behavior of the system in time and space can be predicted and combined with PK/PD models for the drug action mechanism

basic difference between systems biology and systems medicine is that systems biology assumes the data to be useable and correct, while systems medicine ensures the validity of molecular and clinical data sets to interpret the pathways that may contribute to therapeutic possibilities to the adapted healthcare [19, 20].

### 13.2.2 Systems Medicine for Human Diseases and Novel Drugs Research

Systems medicine requires different features to achieve clinical and diagnostic goals [19]. One of the important areas in the systems medicine is the progress of the computational models that help to explain the disease advancement and effects of therapeutic interventions [21]. This is important for the better control of large data sets and also to elucidate wet laboratory in order to develop multifaceted interrelationships among molecular targets.

Systems medicine plays a vital role in drug development whereby drugs are proven to be effective for one condition or ineffective for a different medical condition [22]. Jin et al. performed transcriptome expression analysis before and after the drug administration to observe the off-target effects of drug for signaling pathways. This study recognized a systems-based analytical approach named as Bayesian factor regression model (BFRM) accompanied by cancer signaling bridges (CSB), termed as CSB-BFRM, which is fruitful in the prediction of outcomes of clinical responses arising for Food and Drug Administration (FDA)-approved drugs through validation using three independent cancer models, thereby assuring the accuracy of systems medicine approach [23].

Systems medicine has an impact in recognizing innovative disease networks. The foremost exploration focuses on the connections of models that are influenced by the pathogenesis and are inactive or active in numerous disease conditions. MicroRNA (miRNA) research is one of the typical methodologies to recognize the application of systems medicine. As miRNA controls the transcripts, one miRNA perhaps deregulates the expression of numerous downstream target genes; therefore, it is possible that miRNA can be applied in many clinical conditions probably in a simultaneous manner [24]. Outline of systems medicine is shown in Fig. 13.2.

Systems medicine is paving its way for academics, clinicians, and researchers dealing with experimental research approaches. The probability to investigate an immense data from in silico and experimental approaches offers more understanding into the complex molecular interactions. This assists to the enlightening of unusual dynamic interactions that are vital for medical conditions and therefore serve as clinically significant key molecules for future therapeutics [18, 24].

## 13.3 Computational Approaches

Extensive measurements of somatic mutations in the tumors are possible through high-throughput DNA sequencing technologies. Cancer genomics purposes to find out all the genes related to cancer and their involvements in cancer development. Cancer-driven mutation and pathways can be detected on the basis of biological networks and different computational approaches. They can be classified into (1)

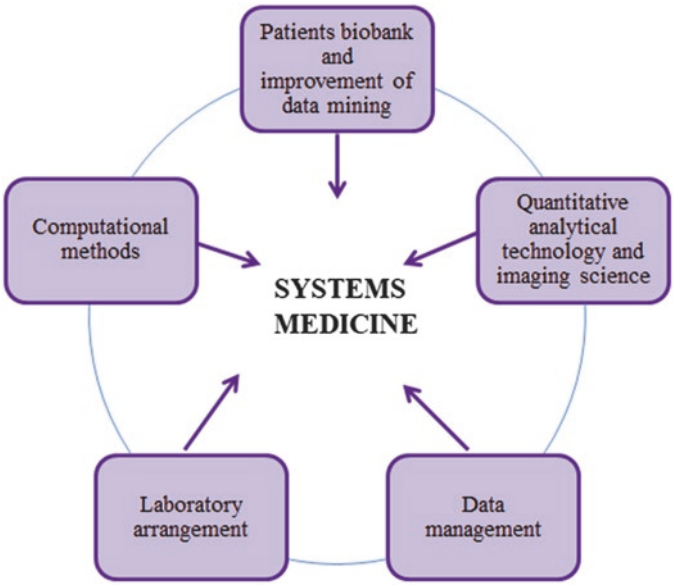


**Fig. 13.2** Outline of the mandatory features for execution of systems medicine methods to modern medical research

functional impact-based approach, (2) network- or pathway-based approach, (3) data integration-based approach, (4) mutation frequency-based approach, and (5) structural genomics-based approach [25]. Here, the approaches and the databases used for the identification of cancer genes and pathways will be focused.

### 13.3.1 Data Resources of Cancer-Related Genes, Networks, and Pathways

Different databases are present that contain information and function about cancer genes. Among them, COSMIC (The Catalogue Of Somatic Mutations In Cancer) is one of the largest databases. It includes mutations from the cancer cell lines and also the whole genome and exome of patients having cancer and hence provides detailed information of somatic mutations [26]. The Cancer Genome Atlas (TCGA) characterizes genomic changes in 33 cancer types which has enhanced the evaluation of genomic changes in cancer genomics. Single base substitutions in TCGA are 2,948,799, among them 1,648,416 are missense variants [27]. The International Cancer Genome Consortium (ICGC) aims to describe the epigenomic, transcriptomic, and genomic profiles of the cancer genomes of 50 different cancer types [28].

cBioportal is a web source of visualizing and investigating cancer genomics data [29]. These annotation databases are helpful to decode the consequences among mutations and protein 3D structures. To identify driver mutations specifically in kinase domain, protein's three-dimensional (3D) structure information is used. In context of the 3D structure, another database is Cancer3D to investigate missense somatic mutations [30]. dSysMap is a resource for mapping the missense mutations through the structurally annotated interactome of human. Recently, for studying function of noncoding somatic mutations, different projects have been initialized as protein-coding human genome is just <2% [31]. These include Encyclopedia of DNA Elements (ENCODE) [32], the functional annotation of the mammalian genome 5 (FANTOM5) [33], and NIH Roadmap Epigenomics [34]. These databases offer comprehensive resources of functional genomics data to describe regulatory role of noncoding mutations. Genotype-Tissue Expression (GTEx) project delivers genetic expression and regulation data for many human tissues. It helps to study the tissue-specific regulatory pathways that are changed by somatic mutations Consortium GT. Human genomics [35]. To study somatic cancer mutation, the Database of Curated Mutations (DoCM) is used. It includes 1276 missense mutations and 1364 variants from 122 cancer subtypes [36]. Another community-edited web source named as Clinical Interpretations of Variants in Cancer (CIViC) is used for discovering different variants in cancer. It includes 1767 variants until February 2018 and enables precision medicine for cancer treatment [37]. Table 13.1 [38] shows all the data resources for cancer-driven mutations.

**Table 13.1** Data resources for the assessment of computational tools for somatic mutation genes and driver mutations in cancer

| Name | Depiction |
|---|---|
| COSMIC | Comprehensive resources of somatic mutations |
| TCGA | Characterize genomic changes in 33 cancer types |
| ICGC | Describe epigenomic, transcriptomic, and genomic profiles of the cancer genomes |
| cBioPortal | Visualization and investigation of cancer genomics data |
| Cancer3D | Functional roles of somatic mutations via protein 3D structure |
| dSysMap | For mapping the missense mutation on the structurally annotated interactome of human |
| ENCODE | Comprehensive resources of functional genomics data |
| NIH Roadmap Epigenomics | Resources of functional genomics data |
| FANTOM | Regulatory role of noncoding mutations |
| GTEx | A resource for the tissue-specific regulation and gene expression |
| DoCM | For somatic cancer mutations |
| CIViC | For variants in cancer |

### 13.3.2 Data Resources for Networks and Pathways

Detailed analysis based on gene networks has been applied to interpret somatic mutations in the cancer [39]. Protein-protein interaction (PPI) and pathway-related databases have been established such as Reactome [40], WikiPathways [41], Pathway Interaction Database (PID) [42], and Pathway Commons [43]. These databases have been widely used to assess the role of variants and somatic mutations [44].

Some important PPI databases include BioGRID [45], HPRD [46], MINT [47], IntAct [48], STRING [49], PINA [50], PhosphoSitePlus [51], Phospho.ELM [52], PTMcode [53], Interactome3D [54], Instruct [55], and 3did [56].

PPI databases provide a network resource of complementary molecular interactions to decipher the consequences of somatic variations in various cancers as they enlist literature-derived and experimental PPIs, 3D structure PPIs, and kinase-substrate-specific phosphorylation events (Fig. 13.3).

## 13.4 Computational Approaches and Methods

Computational methods help in the fastest way to characterize the disease. General approaches used for the investigation of somatic mutations are shown in Fig. 13.4. Through whole genome sequencing, list of mutations leading to cancer can be obtained.

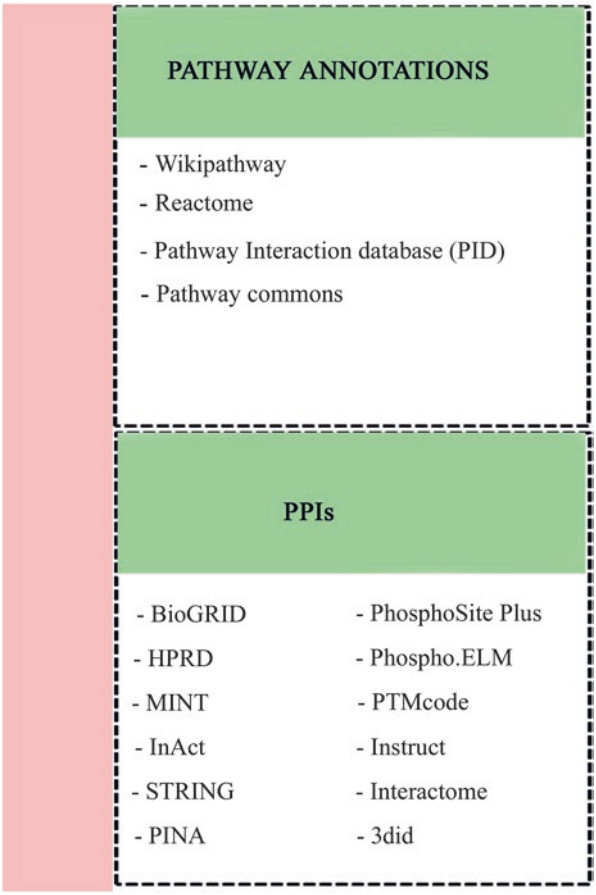


**Fig. 13.3** Data resources for prioritizing driver mutations and pathways in cancer

### 13.4.1 Mutation Frequency-Based Approaches

Significantly mutated genes (SMGs) in the cancer are defined by categorizing the genes that undergo more mutations than those based on the mutation model in a certain cancer type [57]. Table 13.2 [38] summarizes the computational approaches based on the mutation frequency such as Mutational Significant in Cancer (MuSiC). It incorporates the clinical data with sequence-based data to find out the relationship among affected genes, mutations, and pathways [58]. Similarly, ContrastRank compares alleged defective rate of every gene against normal data [59]. As the model with low mutation frequency may lead to false positive results, thus other methods were projected. SMGs based on the gain of function mutation can be identified by OncodriveCLUST [60]. It showed that silent mutations play a vital role in cancer. OncodriveCLUST uses silent mutation as the background. Lawrence et al.

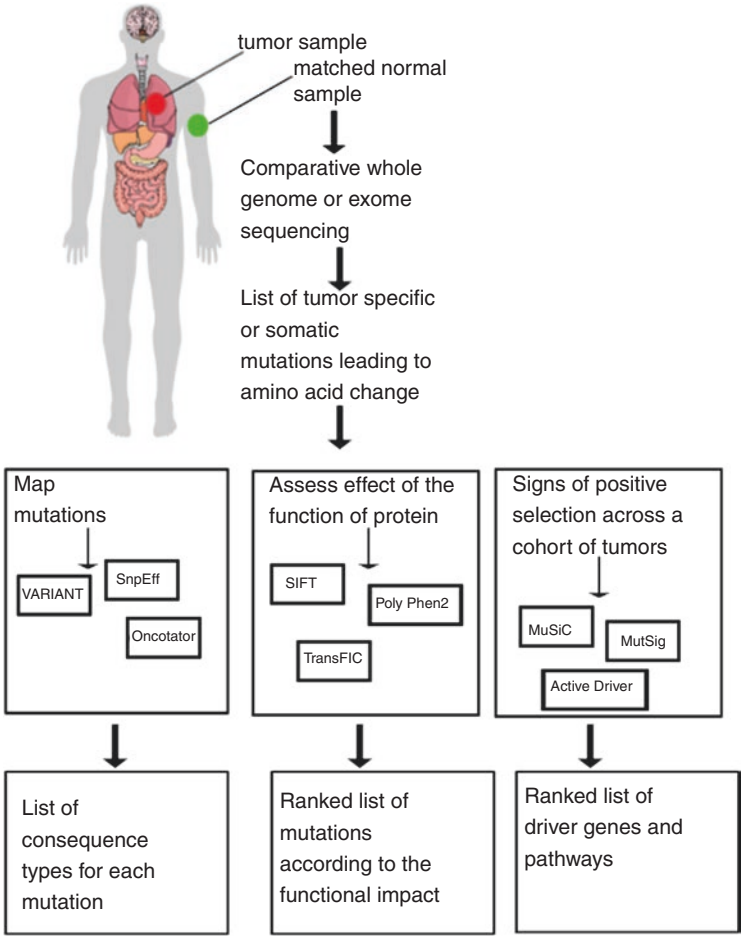


**Fig. 13.4** General approaches used in the routine for the analysis of cancer somatic mutation

established MutSigCV that uses the information of replication timing and gene expression to develop a patient-specific mutation model [57].

### 13.4.2 Functional Impact-Based Approaches

Computational methods offer a fast and an economical way to evaluate the impact of mutations. These methods help the researchers to find the putative mutations that can validate their experimental work. Multiple tools have been developed for the computational approaches. One of the tools is SIFT (The Sorting Intolerant from Tolerant) that finds out the impact of amino acid substitution in protein function. It is based on the extent to which an amino acid is conserved in sequence alignment

**Table 13.2** Summary of tools and computational approaches for the identification of driver mutations and SMGs in cancer genome

| Mutation frequency-based approaches | |
|---|---|
| MuSiC | An approach for determination of the mutational significance in cancer |
| MutSigCV | An integrative approach that corrects for variants using patient-specific mutation frequency and spectrum and gene-specific background mutation model derived from gene expression and replication timing information |
| OncodriveCLUST | Identifying genes with a significant bias toward mutation clustering in specific regions of proteins using silent mutations as a background mutation model |
| ContrastRank | A method based on estimating the assumed defective rate of each gene in tumor against normal samples from the 1000 Genomes Project data |
| *Functional impact-based approaches* | |
| MutationTaster | A tool including evolutionary conservation and splice-site change information for the prediction of functional impacts of DNA sequencing modifications |
| MutationAssessor | Based on evolutionary conservation patterns, it can predict the functional impacts |
| SIFT | A tool that uses protein sequence homology for the prediction of biological effect of missense variations |
| PolyPhen-2 | A tool for the prediction of the functional impacts of protein sequence variants by using three structure-based and eight sequence-based predictive features to build naive Bayes classifiers |
| CHASM and SNVbox | Python programs that use the tumorigenic impact of mutations for cancer-related mutations |
| Condel | A consensus deleteriousness score for evaluating the functional impact of missense mutations |
| OncodriveFM | An approach based on functional impact bias using three well-known methods |
| CanDrA | A tool based on a set of 95 structural and evolutionary features |
| PROVEAN | For the prediction of functional effects of SNV and in-frame insertions and deletions |
| FATHMM | A tool based on Hidden Markov model-based tool for functional analysis of driver mutations |
| CRAVAT | A web-based toolkit for arranging missense mutations related to tumorigenesis |
| *Data integration-based approaches* | |
| MAXDRIVER | An approach that uses the data from copy number variant regions of cancer genomes for the prediction of SMGs |
| CONEXIC | A computational framework that assimilates copy number variants and gene expression changes for prioritizing SMGs |
| CAERUS | An approach for the prediction of SMGs using structural information of proteins, protein networks, gene expression, and mutation data |
| Helios | For prediction of SMGs by the integration of functional and genomic RNAi screening data |
| OncoIMPACT | A framework based on phenotypic impacts for highlighting SMGs |

(continued)

**Table 13.2** (continued)

| Mutation frequency-based approaches | |
|---|---|
| OncodriverROLE | An approach that classifies SMGs into LoF and GoF |
| DOTS-Finder | A tool based on functional and frequency for predicting SMGs in cancer |
| *Structural genomics-based approaches* | |
| ActiveDriver | For prediction of SMGs having driver mutations significantly changing phosphorylation sites of proteins |
| iPAC | For the prediction of SMGs by using protein 3D structure |
| MSEA | For the prediction of SMGs based on mutation patterns on domains of protein |
| CanBind | For the prediction of SMGs using the information on the binding site of protein–ligand |
| *Network or pathway-based approaches* | |
| PARADIGM-SHIFT | It uses belief-propagation algorithm for ordering downstream pathways by a mutation in cancer |
| PARADIGM | By incorporation of patient-specific genetic data, it detects consistent pathways in cancers |
| DriverNet | By estimating the effect on mRNA expression networks, it identifies driver mutations |
| Personalized pathway enrichment map | From individual genome, it identifies alleged cancer genes and pathways |
| NBS | An approach for stratifying tumor mutations |
| TieDIE | An approach for identification of cancer-mutated subnetworks |
| DawnRank | On the basis of PageRank algorithm, it prioritizes SMGs in a single patient |
| HotNet2 | For the detection of mutated subnetworks in cancer, an algorithm is used to overcome the limitations of existing single-gene and network approaches |
| VarWalker | A novel approach for prioritizing SMGs |

derived from the closely related sequences [61, 62]. SIFT can characterize impact of missense mutations. Another software named as Polymorphism Phenotyping v2 (PolyPhen-2) is used with SIFT for better results. It predicts the impact of the variants by three structure-based and eight sequence-based features [63]. Another web server, MutationAssessor, uses a novel functional impact score for the characterization of residual mutation. To define the evolutionary conservation patterns, which are taken from aligned families and subfamilies, it uses combinatorial entropy formalism [64]. The three methods mentioned above are useful for nonsynonymous SNVs only. Multiple methods incorporate domain information to predict the functional impact of SNVs. One of them is OncodriveFM. It identifies low recurrent candidate SMGs by utilizing the features of SIFT, MutationAssessor, and PolyPhen-2 [65]. For rapid evaluation of DNA sequence, alteration that is involved in causing the disease can be assessed from MutationTaster. It uses the information from splice-site changes, conservation, and loss of protein features [66]. For somatic missense prediction, CHASM is used. It uses a Random Faster classifier trained with 49 predictive features [67]. Another software based on Hidden Markov model, known

as FATHMM, helps in finding the cancer-associated mutations. It differentiates passenger mutations from the amino acid substitutions associated with the cancer. This is achieved by integrating homologous sequence alignment and information of conserved protein domains [68].

CRAVAT toolkit is used to highlight SMGs and mutations using SNVbox and CHASM [69]. Machine learning-based tool CanDrA is based on supporting vector machine (SVM) that incorporates 95 evolutionary and structural features for ranking SMGs [70].

Despite the existence of multiple strategies, there are some limitations of these tools including lack of standard and positive results and selection of nonfunctional mutations.

### 13.4.3 Data Integration-Based Approach

Cancer data include transcriptome, somatic mutation, proteomics, methylation, and profiles of a tumor and matched normal tissues. It enables the investigators to investigate SMGs and mutations for precision medicine [71]. Data integration-based approaches include Driver Oncogene and Tumor Suppressor (DOTS)-Finder. It categorizes SMGs in cancer by integrating three features of a mutated gene: (1) mutation pattern, (2) mutation frequency, and (3) effect on the gene product's function due to the mutation [72]. It can also predict SMGs specific to oncogenes or tumor suppressor genes. Another unique pipeline SVMerge detects the breakpoints and structural variants by local assembly information and structural variant algorithms [73].

In this regard a favorable direction is to develop an approach that uses the structural variant data like CNVs to rank the SMGs and driver mutations. Driver mutations related to cancer can be identified by CONEXIC. It is done by integrating the CNVs and the genetic expression from tumor-normal samples [74]. They have also developed an algorithm, called Helios, that identifies SMGs within the amplified DNA regions by incorporating cancer genomics data into functional RNA interference (RNAi) data [75]. Helios can assess the potential drivers without a previous genes list.

MAXDRIVER detects alleged SMGs by optimization strategies to build a heterogeneous network by integrating a fused gene functional similarity network with an already existing gene-cancer network [76]. A machine-based learning approach is the OncodriverROLE that categorizes SMGs into activated (Act) and LoF gene [77], although it is a major task for models based on machine learning. A data integration framework OncoIMPACT is based on the phenotypic impacts of patients and forecasts patient-specific SMGs [78].

### 13.4.4 Structural Genomics-Based Approach

With the advancement in technologies like X-ray crystallography and nuclear magnetic resonance, 3D structures have been generated that are available in different

databases like Protein Data Bank (PDB) [79]. In recent years, multiple tools have been developed that require either structure or sequence, because at the structural level, mutations are related with the diseases or drug targets. MSEA (mutation set enrichment analysis) is used to predict alleged SMGs. It is employed using two unique modules (MESA-clust and MESA-domain). MESA-clust is used to screen hotspot regions of mutations by scanning the genomic regions, while MESA-domain is based on the hotspot mutational patterns of protein [80]. Chang et al. developed a network having global kinase-substrate interaction. This network contains 1961 substrates having 36,576 sites for phosphorylation and 7346 pairs connecting 379 kinases [81]. Another approach, ActiveDriver [82], is based on the hypothesis that the cancer-driven mutations may alter the phosphorylation sites of the protein [83]. It analyzes missense point mutations and uses all the phosphorylation sites given in the literature as a mixture training set. A computational pipeline is based on protein pocket to study the functional concerns of somatic mutations in the cancer [84]. Those regions where small molecules and drugs binding occur are known as protein pockets. The mutations lying at these sites may alter the function of protein leading to cancer. SGDriver is based on the relationship among protein 3D structures and somatic mutations to delineate SMG products [85]. SGDriver helps to find out the druggable mutations that can be used in the upcoming field of cancer precision medicine. CanBind is a tool to rank the SMGs that contains the mutations by altering their peptide binding sites or nucleic acids. Identification of Protein Amino acid Clustering iPAC is another algorithm; it prioritizes nonrandom somatic mutations present in the proteins using the 3D structure of a protein [86, 87]. eDriver is another tool to characterize SMGs based on the internal division of somatic missense mutations between protein domains [88]. The development of new tools and approaches will provide exceptional prospects for the clinical applications of cancer genomics data.

### 13.4.5 Network- or Pathway-Based Approach

Various molecular structures of the cell form a dynamic network. Any genetic change in molecular network frame can cause disturbance in the pathway [89]. Large amount of cancer genomics data obtained from the NGS helps to understand the network-level studies of tumor initiation and progression. As cancer is an intricate disease having changes at the network level, hence there is a dire need to characterize the SMGs and driver mutations. A unique method called PARADIGM detects these pathways by incorporating specific genetic data of the patient. PARADIGM-SHIFT includes downstream pathways which are changed due to mutations by incorporation of gene expression, somatic mutations, and CNVs using a belief-propagation algorithm [90]. It identifies potential functional effects as well such as gain of function (GoF) and loss of function (LoF). TieDIE is based on the network diffusion approach. It is used for the prediction of gene expression changes due to genomic alteration [91]. It identifies a cancer-specific subnetwork by the incorporation of transcriptomic and genomics data into networks originated by

PPIs. The downstream transcriptional alterations due to somatic variations are also recognized. DriverNet is a computational network to recognize the mutations by their effects on mRNA expression network [92]. It identifies rare mutations that mediate oncogenic networks. DawnRank is a computational approach to characterize SMGs on an individual patient using PageRank algorithms [93]. The first personalized tool to rank the SMGs by somatic variation is VarWalker. It uses the somatic variation information from the genome and then adjusts gene length by resampling the mutations. It includes cancer genomics data on a large scale using random walk with restart algorithm [94].

Network-based stratification (NBS) is a unique approach based on network. It stratifies cancer subtypes on the basis of somatic mutation profiles presented in an individual tumor [95]. On the basis of genome-scale interaction network, HotNet identifies mutated pathways in cancer [96]. HotNet2 has been developed by the same group for the detection of subnetworks having mutation. It is done by the insulated heat-diffusing process [96]. They recognized 16 considerably mutated subnetworks that include well-known cancer signaling pathways during pan-cancer analysis to recognize the genes that are occasionally mutated in pan-cancer data sets and in individual cancer data sets. (Pan-cancer analysis revealed that some tumors were more likely to be molecularly and genetically the same due to the types of their rising cells instead of the origin of tissue site.) These approaches are successful, but they have some limitations as well, as the current PPI networks cover only 20–30% pairwise PPIs in humans [97]. This shows that current human interactome may be incomplete [98]. Many structural variants, gene expression and methylation patterns, and noncoding variants are not supposed in the abovementioned approach. Another limitation is that the pathways are sometimes prone to error because they are generated on the computational or experimental data, which are always mixed on the condition specificity. Thus development of an integrative framework to improve human interactome knowledge may offer a complete collection of mutated pathways or networks in the cancer.

In Table 13.2 and in Fig. 13.5, all computational approaches used for the mutational analysis and data resources are shown.

As the technology fastens, tool development for the calculations, measurements, assessment, and integration of data is becoming important [99]. Table 13.2 enlists many online databases that are used for storage of genomic-scale data, regulatory sequence [100], and proteomic analysis [101]. These databases provide the data by which cancer models can be evaluated. As the challenges remain, development of more accurate and biologically powerful in silico tools for representation of human cancer is needed. The general resources and databases used in distributing large amount of data are shown in Table 13.3.

## 13.5 Precision Medicine

The concept in precision medicine is based on the lifestyle, environment, and genes of a person. With the advancement in the genetics, we have gained the opportunity to make the personalized care of a patient into reality. Precision medicine for breast

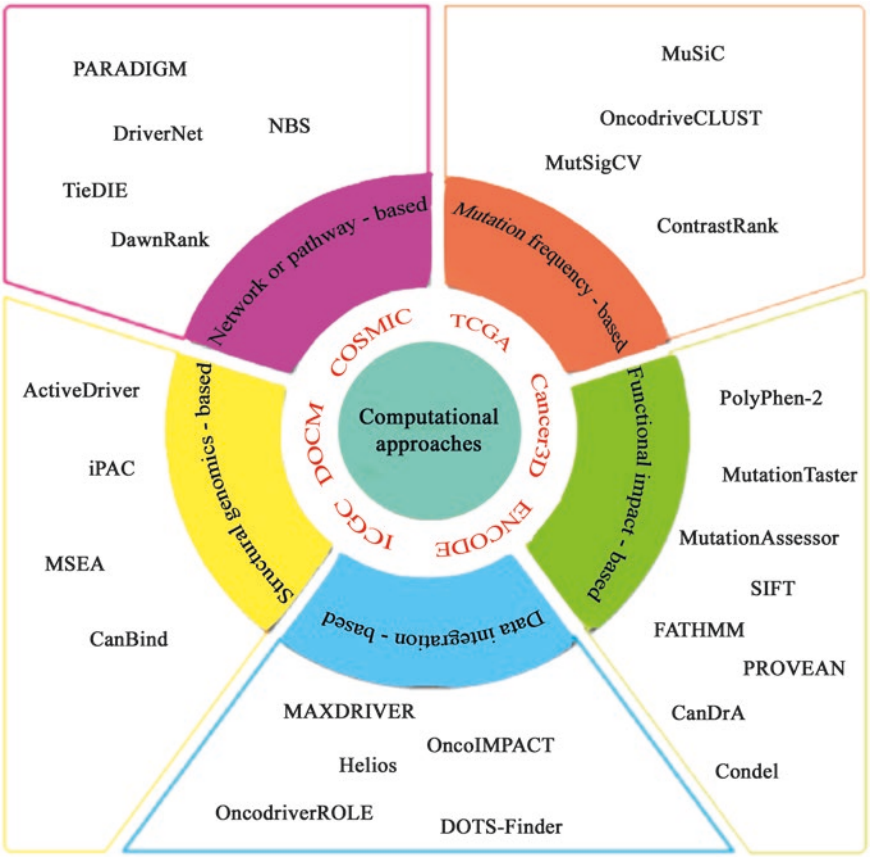


**Fig. 13.5** A summarize form of computational approaches used in the cancer mutation analysis

cancer is the most tempting area, but still it is facing a lot of challenges. Other measures may help in early detection of breast cancer like monitoring of circulating tumor DNA and ultradeep sequencing.

### 13.5.1 Precision Medicine Tools

Identification of genomic changes in patients having the breast cancer helps to adopt the therapy. With the passage of time, different tools are serving for the therapeutic approach; for example, immunohistochemistry was lately used to stratify breast cancer patients with the presence of biomarkers. Now it is used to determine HER2 and ER [125]. To find the copy number, fluorescence in situ hybridization (FISH) is used. DNA array [126], RT PCR [127], or NanoString Technologies [128] is widely used for gene expression quantification. These assays are employed in the early stage detection of breast cancers according to their risk of reversion. NGS is also

**Table 13.3** General resources and databases for in silico analysis of cancer

| Resource | Database | References |
|---|---|---|
| Genome sequence data | Ensemble | Flicek et al. [102] |
| | UCSC genome browser | Karolchik et al. [103] |
| *Genome annotation data* | | |
| Genetic elements | Entrez gene | Maglott et al. [104] |
| | Gene ontology annotation database | Camon et al. [105] |
| | Universal protein knowledge base | Apweiler et al. [106] |
| | Genome reviews | Sterk et al. [107] |
| Biochemical pathways and functional associations | Kyoto encyclopedia of genes and genomes | Ogata et al. [108] |
| | Gene ontology | Ashburner et al. [109] |
| | The SEED | DeJongh et al. [110] |
| | MetaCyc | Krieger et al. [111] |
| | BioCyc | Karp et al. [112] |
| | TransportDB | Ren et al. [113] |
| Regulatory sequences | Eukaryotic promoter database | Cavin Périer et al. [100] |
| | Transcriptional regulatory element database | Zhao et al. [114] |
| Model, model parameter repositories | Kinetic data of biomolecular interactions database | Ji et al. [115] |
| | BioModels database | Le Novère et al. [116] |
| | Database of quantitative cellular signaling | Sivakumaran et al. [117] |
| Protein interaction networks | Database of interacting proteins | Xenarios et al. [118] |
| | Molecular INTeraction database | Zanzoni et al. [119] |
| | Mammalian protein-protein interaction database | Pagel et al. [120] |
| *High-throughput genome-scale data* | | |
| Transcriptomics | Gene expression omnibus | Edgar et al. [121] |
| | Stanford microarray database | Sherlock et al. [122] |
| Proteomics | Proteomics identifications database | Martens et al. [101] |
| Visualization and data management software packages | Cytoscape | Shannon et al. [123] |
| | The Gaggle | Shannon et al. [124] |

used for the identification of dominant mutations in multigene panel. For the detection of minor sub-clonal alterations, ultradeep sequencing can be used. Nucleic acid detection as well as protein expression pattern is required for the comprehensive molecular profile of the tumors.

### 13.5.2 Limitations of Precision Medicine

Although there are multiple applications and high-throughput technologies, still many limitations and several challenges are needed to be addressed. A few of them are described below.

#### 13.5.2.1 Logistical and Operational Challenges

- It is very challenging to complete drug testing trials in genomic segments, although these variations are rare and still randomized clinical trials are needed for the approval.
- Genomic results for a certain amount of patients cannot be delivered as biopsy is not achievable for all the patients. Previously known DNA alterations are not enough to explain the progression of cancer in large amount of patients.
- Development of drug and its access is limited due to lesser amount of patients and locations. Genomic tests are very expensive and unaffordable as a private company runs those genomic tests.

#### 13.5.2.2 Scientific Challenges

- Response rates are very low, as multiple pathways are activated resulting in the failure to recognize oncogenic driver.
- Due to the pressure of treatment, additional genomic changes may occur, causing secondary resistance [129].

## 13.6 Genomic Medicine

Genomic medicine uses the genetic information of an individual as part of his care. It helps to predict disease risk and plots disease course. Genomic medicine makes the plan management according to the need of the patient [130]. The technologies such as high-throughput sequencing and analytical tools help to analyze thousands of molecules simultaneously. Together with computational biology, we can interpret large amount of data sets obtained. The demand for molecular characterization of the disease has been increased to identify the markers for prognosis by the introduction of targeted therapy. This also assists in developing new therapies [131]. Such analyses will also help in early cancer detection and better treatment [132].

### 13.6.1 Genomic Sequencing for Assessment of Disease

In personalized medicine, NGS has provided us with several promising applications. Genome sequencing may also provide important assistance for reproductive health. This includes prescreening of mothers for mutations related to metabolic and other disorders [133]. Exome sequencing also offers molecular-based diagnosis as it identifies the novel mutation.

The applications of genomic medicine are as follows:

1. Inspection of difference among healthy individuals
2. Disease hindrance
3. Understanding disease risk, susceptibility, and etiology
4. Diagnosis of challenging cases with indecisive results for clinical parameters
5. Classification of accurate disease based on molecular signature
6. Early diagnosis to modify disease course
7. Identification of new mutations related to disease
8. Development of new targeted therapies
9. Personal drug-related profile identification
10. Patients selection for clinical trials
11. Monitoring disease status
12. Evolution of tumor in response to treatment
13. Health management

Furthermore, risk assessment for diseases like diabetes, cancer, and hypertension is economically efficient. It will significantly decrease the treatment problems and may be followed up for prolonged time period [134]. In Fig. 13.6, multistep process is shown.

### 13.6.2 Genomics Databases

Many genome-wide studies have been applied for the analysis of single nucleotide polymorphisms (SNPs) to examine the genetic variants in different individuals and their effects and its relation with disease risk. In 2005, age-related macular degeneration was investigated [135]. Since then, almost 4000 more associations of SNPs with the disease have been identified [136]. Several international projects have been designed on the oncology frontier to enlist somatic alterations at different levels through exome sequence analysis, mRNA and microRNA (miRNA) production, DNA copy numbers, and promoter methylation. These projects include the Cancer Genome Atlas (http://cancergenome.nih.gov/) [137], the Cancer Genome Project [138], and Hudson et al. [28]. Furthermore, NIH has initiated extensive genomic variation analyses in different diseases by launching various initiatives. Overall, collection of large amount of data at different levels holds a great promise to understand disease management [133]. There are several databases that collect the data to gain a meaningful conclusion.

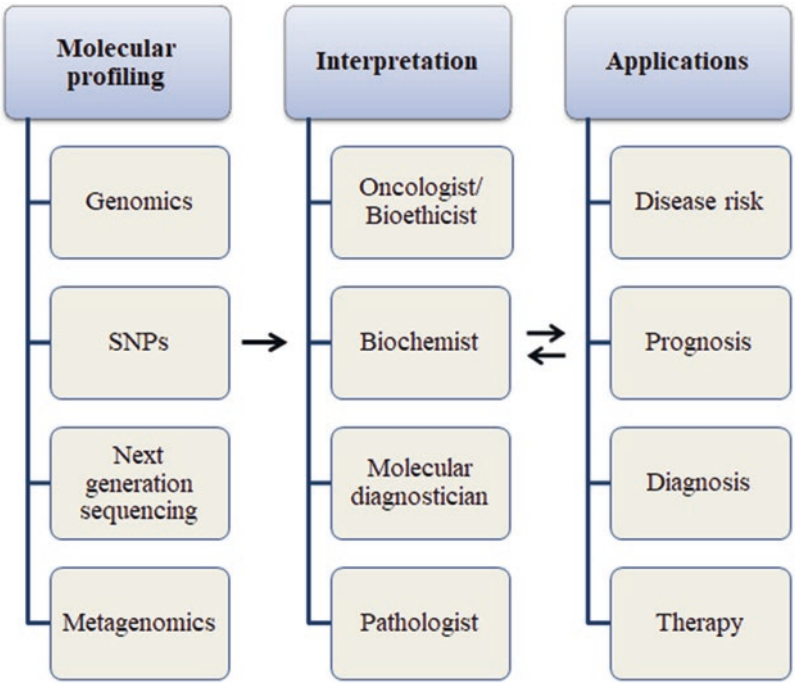


**Fig. 13.6** A multistep process for translating genomics to clinical data

Examples of such databases are i2b2 (Informatics for Integrating Biology and the Bedside; https://www.i2b2.org) [139] and locus-specific mutation databases, such as the Human Gene Mutation Database or HGMD (http://www.hgmd.cf.ac.uk/ac/index.php) [140]. Hence data taken from NGS must be inferred in the perspective of environmental conditions and clinical variables for better results.

### 13.6.3 Monitoring the Personal Genome

Integrative personal omics profiling (iPOP) is a new approach for monitoring personal genome as it combines metabolomics, genomic, proteomic, transcriptomic, and autoantibody profiles of the same person to follow genomic and transcriptomic composition over long periods. By the connection of genetic information with dynamic "omics" activities, it can evaluate disease state and healthy state. These profiles associated with different states are integrated in this approach at multiple time points. An extensive database may be generated with the profiles from more individuals having different kinds of diseases. Such databases might be useful in the monitoring, diagnosis, and disease treatment [141].

### 13.6.4 Potential Challenges of Genomic Medicine

Although a lot of development has been made, still there are some challenges in the genomic medicine; few of them are given here:

- It is difficult to interpret data and extract actionable items.
- Rules must be set in implementation of new molecular tests.
- Cost-effectiveness. An important apprehension in molecular testing.
- Patient heterogeneity that occurs with the same cancer type and ethnic variation while interpreting genomics data must be addressed carefully.
- "Test accuracy" should improve with time.
- There also is a huge risk of incidental findings and false-positive results.
- Training and teamwork efforts are also needed [142].

## 13.7 Mathematical Models

Mathematical models allow the researchers and investigators in intricating processes that are connected to each other and how their disturbance leads to the disease development. It also helps to analyze system perturbations systematically and to develop hypothesis for the development of new tests for experiments. Ultimately, new therapeutic targets can be evaluated. Models that describe biological system are very complex to handle manually that is why they are handled numerically. One of the biggest advantages of the mathematical model for the biological systems is computer simulations. These simulations have a lot of benefits. Firstly, a comprehensive molecular scenario can be seen by looking at the discrepancies between the behavior of system projected by mathematical modeling and its actual behavior calculated in experiments. Secondly, with the help of mathematical modeling system, various perturbations can be seen, for example, after drug administration and developmental signals, etc. Thirdly, mathematical simulations are not bound like wet experiments; different experimental conditions can quickly be investigated by computer simulations [143].

### 13.7.1 Mathematical Equations for Biological Systems Behaviors Modeling

Understanding the biological system is the first step for modeling as different kinds of mathematical frameworks have been developed to model various biological systems. It is important to understand the biological process for selecting the optimal modeling approach because for modeling of different biological systems, diverse mathematical frameworks have been developed. For instance, dynamic processes govern different cellular systems so that the cell adapts its environmental changes. For the description of time-dependent phenomena, it is vital to select mathematical equations that can capture the dynamic effects. Modeling of metabolic processes is

essential for a living organism. It provides the energy to the cell by delivering building blocks for the large molecules. Biological research has been dedicated to metabolism for many years, and still full pathways are not known. A main factor is metabolic flux in any metabolic study, that is, conversion rate of metabolites together with a metabolic pathway.

Modeling of signaling and regulatory pathways functions as the central control machinery of a cell. It firmly regulates responses of the cell to the stimuli. These pathways involve the signal transmission from cell membrane into the nucleus of the cell. Pathways are mainly triggered by binding of certain extracellular biomolecules to the receptor as a result; the receptor's 3D structure may be changed. Modeling of comparatively simpler signaling networks revealed that signal transmission from the cell shows unexpected behaviors, such as periodic enhancement patterns of the initial signals [144].

## 13.8 Conclusion

NGS have assisted researchers to produce large amount of somatic mutations and cancer genomics data in rare and common cancer types. Genetic alterations containing small insertions or deletions, single nucleotide variants, large chromosomal rearrangements; gene fusions are cause of causing cancers. Many computational tools have been developed for pinpointing the cancer genes and driver mutations from millions of somatic cancer mutations. The chapter focused on computational methods for the prediction of mutations based on their structure, analysis of missense mutations in the 3D protein structure, and its effects on stability and interactions. Albeit cancer genomics is still in its beginning, the exceptional production of cancer genomics data assured the better prediction of novel cancer genes. With the increase in number of tumor samples, these computational methods and approaches helped in interpretation of tumor heterogeneity. It facilitated the identification of cancer-driven mutations and delineation of dysregulated pathways which can be targeted by drugs through precision and genomic medicine.

## References

1. Breitling R (2010) What is systems biology? Front Physiol 1:9. https://doi.org/10.3389/fphys.2010.00009
2. Palsson BO, Abrams M (2011) Systems biology: simulation of dynamic network states. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511736179
3. Tavassoly I, Goldfarb J, Iyengar R (2018) Systems biology primer: the basic methods and approaches. Essays Biochem 62:487–500. https://doi.org/10.1042/EBC20180003
4. Chergui M (2009) Physical biology from atoms to medicine edited by Ahmed Zewail. Angew Chem Int Ed 48:3014–3016. https://doi.org/10.1002/anie.200900611
5. Longo G, Montévil M (2014) Perspectives on organisms. Springer, Berlin. https://doi.org/10.1007/978-3-642-35938-5

6. Hood L, Heath JR, Phelps ME, Lin B (2004) Systems biology and new technologies enable predictive and preventative medicine. Science 306:640–643. https://doi.org/10.1126/science.1104635
7. Barillot E (2012) Computational systems biology of cancer. CRC Press, Boca Raton. https://doi.org/10.1201/b12677
8. Werner HMJ, Mills GB, Ram PT (2014) Cancer systems biology: a peek into the future of patient care? Nat Rev Clin Oncol 11:167–176. https://doi.org/10.1038/nrclinonc.2014.6
9. Gentles AJ, Gallahan D (2011) Systems biology: confronting the complexity of cancer. Cancer Res 71:5961–5964. https://doi.org/10.1158/0008-5472.CAN-11-1569
10. Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. Genomics 92:255–264. https://doi.org/10.1016/j.ygeno.2008.07.001
11. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63. https://doi.org/10.1038/nrg2484
12. Rual J-F, Venkatesan K, Hao T et al (2005) Towards a proteome-scale map of the human protein-protein interaction network. Nature 437:1173–1178. https://doi.org/10.1038/nature04209
13. German JB, Hammock BD, Watkins SM (2005) Metabolomics: building on a century of biochemistry to guide human health. Metabolomics 1:3–9. https://doi.org/10.1007/s11306-005-1102-8
14. Tyson JJ, Novák B (2010) Functional motifs in biochemical reaction networks. Annu Rev Phys Chem 61:219–240. https://doi.org/10.1146/annurev.physchem.012809.103457
15. Kafri R, Levy J, Ginzberg MB et al (2013) Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle. Nature 494:480–483. https://doi.org/10.1038/nature11897
16. Iyengar R, Zhao S, Chung S-W et al (2012) Merging systems biology with pharmacodynamics. Sci Transl Med 4:126ps7. https://doi.org/10.1126/scitranslmed.3003563
17. Federoff HJ, Gostin LO (2009) Evolving from reductionism to holism: is there a future for systems medicine? JAMA 302:994–996. https://doi.org/10.1001/jama.2009.1264
18. Ayers D, Day PJ (2015) Systems medicine: the application of systems biology approaches for modern medical research and drug development. Mol Biol Int 2015:698169. https://doi.org/10.1155/2015/698169
19. Cardinal-Fernández P, Nin N, Ruíz-Cabello J, Lorente JA (2014) Systems medicine: a new approach to clinical practice. Arch Bronconeumol 50:444–451. https://doi.org/10.1016/j.arbr.2014.09.001
20. Tillmann T, Gibson AR, Scott G et al (2015) Systems medicine 2.0: potential benefits of combining electronic health care records with systems science models. J Med Internet Res 17:e64. https://doi.org/10.2196/jmir.3082
21. Costa J (2008) Systems medicine in oncology. Nat Rev Clin Oncol 5:117–117. https://doi.org/10.1038/ncponc1070
22. Jin G, Wong STC (2014) Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines. Drug Discov Today 19:637–644. https://doi.org/10.1016/j.drudis.2013.11.005
23. Jin G, Fu C, Zhao H et al (2012) A novel method of transcriptional response analysis to facilitate drug repositioning for cancer therapy. Cancer Res 72:33–44. https://doi.org/10.1158/0008-5472.CAN-11-2333
24. Mestdagh P, Lefever S, Pattyn F et al (2011) The microRNA body map: dissecting microRNA function through integrative genomics. Nucleic Acids Res 39:e136. https://doi.org/10.1093/nar/gkr646
25. Dimitrakopoulos CM, Beerenwinkel N (2017) Computational approaches for the identification of cancer genes and pathways. Wiley Interdiscip Rev Syst Biol Med 9:e1364. https://doi.org/10.1002/wsbm.1364
26. Forbes SA, Bindal N, Bamford S et al (2011) COSMIC: mining complete cancer genomes in the catalogue of somatic mutations in cancer. Nucleic Acids Res 39:D945–D950. https://doi.org/10.1093/nar/gkq929

27. Weinstein JN, Collisson EA, Cancer Genome Atlas Research Network et al (2013) The Cancer Genome Atlas pan-cancer analysis project. Nat Genet 45:1113–1120. https://doi.org/10.1038/ng.2764
28. Hudson TJ, Anderson W, International Cancer Genome Consortium et al (2010) International network of cancer genome projects. Nature 464:993–998. https://doi.org/10.1038/nature08987
29. Gao J, Aksoy BA, Dogrusoz U et al (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 6:11. https://doi.org/10.1126/scisignal.2004088
30. Porta-Pardo E, Hrabe T, Godzik A (2015) Cancer3D: understanding cancer mutations through protein structures. Nucleic Acids Res 43:D968–D973. https://doi.org/10.1093/nar/gku1140
31. Venter JC, Adams MD, Myers EW et al (2001) The sequence of the human genome. Science 291:1304–1351. https://doi.org/10.1126/science.1058040
32. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74. https://doi.org/10.1038/nature11247
33. Andersson R, Gebhard C, Miguel-Escalada I et al (2014) An atlas of active enhancers across human cell types and tissues. Nature 507:455–461. https://doi.org/10.1038/nature12787
34. Kundaje A, Meuleman W, Roadmap Epigenomics Consortium et al (2015) Integrative analysis of 111 reference human epigenomes. Nature 518:317–330. https://doi.org/10.1038/nature14248
35. GTEx C (2015) The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348:648–660
36. Ainscough BJ, Griffith M, Coffman AC et al (2016) DoCM: a database of curated mutations in cancer. Nat Methods 13:806–807. https://doi.org/10.1038/nmeth.4000
37. Griffith M, Spies NC, Krysiak K et al (2017) CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nat Genet 49:170–174. https://doi.org/10.1038/ng.3774
38. Cheng F, Zhao J, Zhao Z (2016) Advances in computational approaches for prioritizing driver mutations and significantly mutated genes in cancer genomes. Brief Bioinform 17:642–656. https://doi.org/10.1093/bib/bbv068
39. Cheng F, Jia P, Wang Q et al (2014) Studying tumorigenesis through network evolution and somatic mutational perturbations in the cancer interactome. Mol Biol Evol 31:2156–2169. https://doi.org/10.1093/molbev/msu167
40. Croft D, Mundo AF, Haw R et al (2014) The reactome pathway knowledgebase. Nucleic Acids Res 42:D472–D477. https://doi.org/10.1093/nar/gkt1102
41. Kelder T, van Iersel MP, Hanspers K et al (2012) WikiPathways: building research communities on biological pathways. Nucleic Acids Res 40:D1301–D1307. https://doi.org/10.1093/nar/gkr1074
42. Schaefer CF, Anthony K, Krupa S et al (2009) PID: the pathway interaction database. Nucleic Acids Res 37:D674–D679. https://doi.org/10.1093/nar/gkn653
43. Cerami EG, Gross BE, Demir E et al (2011) Pathway Commons, a web resource for biological pathway data. Nucleic Acids Res 39:D685–D690. https://doi.org/10.1093/nar/gkq1039
44. Leiserson MDM, Vandin F, Wu H-T et al (2015) Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. Nat Genet 47:106–114. https://doi.org/10.1038/ng.3168
45. Chatr-Aryamontri A, Breitkreutz B-J, Heinicke S et al (2013) The BioGRID interaction database: 2013 update. Nucleic Acids Res 41:D816–D823. https://doi.org/10.1093/nar/gks1158
46. Keshava Prasad TS, Goel R, Kandasamy K et al (2009) Human protein reference database--2009 update. Nucleic Acids Res 37:D767–D772. https://doi.org/10.1093/nar/gkn892
47. Ceol A, Chatr Aryamontri A, Licata L et al (2010) MINT, the molecular interaction database: 2009 update. Nucleic Acids Res 38:D532–D539. https://doi.org/10.1093/nar/gkp983
48. Kerrien S, Aranda B, Breuza L et al (2012) The IntAct molecular interaction database in 2012. Nucleic Acids Res 40:D841–D846. https://doi.org/10.1093/nar/gkr1088

49. Franceschini A, Szklarczyk D, Frankild S et al (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41:D808–D815. https://doi.org/10.1093/nar/gks1094
50. Cowley MJ, Pinese M, Kassahn KS et al (2012) PINA v2.0: mining interactome modules. Nucleic Acids Res 40:D862–D865. https://doi.org/10.1093/nar/gkr967
51. Hornbeck PV, Kornhauser JM, Tkachev S et al (2012) PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res 40:D261–D270. https://doi.org/10.1093/nar/gkr1122
52. Diella F, Cameron S, Gemünd C et al (2004) Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. BMC Bioinf 5:79. https://doi.org/10.1186/1471-2105-5-79
53. Minguez P, Letunic I, Parca L, Bork P (2013) PTMcode: a database of known and predicted functional associations between post-translational modifications in proteins. Nucleic Acids Res 41:D306–D311. https://doi.org/10.1093/nar/gks1230
54. Mosca R, Céol A, Aloy P (2013) Interactome3D: adding structural details to protein networks. Nat Methods 10:47–53. https://doi.org/10.1038/nmeth.2289
55. Meyer MJ, Das J, Wang X, Yu H (2013) INstruct: a database of high-quality 3D structurally resolved protein interactome networks. Bioinformatics 29:1577–1579. https://doi.org/10.1093/bioinformatics/btt181
56. Mosca R, Céol A, Stein A et al (2014) 3did: a catalog of domain-based interactions of known three-dimensional structure. Nucleic Acids Res 42:D374–D379. https://doi.org/10.1093/nar/gkt887
57. Lawrence MS, Stojanov P, Polak P et al (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 499:214–218. https://doi.org/10.1038/nature12213
58. Dees ND, Zhang Q, Kandoth C et al (2012) MuSiC: identifying mutational significance in cancer genomes. Genome Res 22:1589–1598. https://doi.org/10.1101/gr.134635.111
59. Tian R, Basu MK, Capriotti E (2014) ContrastRank: a new method for ranking putative cancer driver genes and classification of tumor samples. Bioinformatics 30:i572–i578. https://doi.org/10.1093/bioinformatics/btu466
60. Tamborero D, Gonzalez-Perez A, Lopez-Bigas N (2013) OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. Bioinformatics 29:2238–2244. https://doi.org/10.1093/bioinformatics/btt395
61. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4:1073–1081. https://doi.org/10.1038/nprot.2009.86
62. Sim N-L, Kumar P, Hu J et al (2012) SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res 40:W452–W457. https://doi.org/10.1093/nar/gks539
63. Adzhubei IA, Schmidt S, Peshkin L et al (2010) A method and server for predicting damaging missense mutations. Nat Methods 7:248–249. https://doi.org/10.1038/nmeth0410-248
64. Reva B, Antipin Y, Sander C (2011) Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res 39:e118. https://doi.org/10.1093/nar/gkr407
65. Gonzalez-Perez A, Lopez-Bigas N (2012) Functional impact bias reveals cancer drivers. Nucleic Acids Res 40:e169. https://doi.org/10.1093/nar/gks743
66. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D (2010) MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 7:575–576. https://doi.org/10.1038/nmeth0810-575
67. Wong WC, Kim D, Carter H et al (2011) CHASM and SNVBox: toolkit for detecting biologically important single nucleotide mutations in cancer. Bioinformatics 27:2147–2148. https://doi.org/10.1093/bioinformatics/btr357

68. Shihab HA, Gough J, Cooper DN et al (2013) Predicting the functional consequences of cancer-associated amino acid substitutions. Bioinformatics 29:1504–1510. https://doi.org/10.1093/bioinformatics/btt182
69. Douville C, Carter H, Kim R et al (2013) CRAVAT: cancer-related analysis of variants toolkit. Bioinformatics 29:647–648. https://doi.org/10.1093/bioinformatics/btt017
70. Mao Y, Chen H, Liang H et al (2013) CanDrA: cancer-specific driver missense mutation annotation with optimized features. PLoS One 8:e77945. https://doi.org/10.1371/journal.pone.0077945
71. Brunak S, De La Vega FM, Rätsch G, Stuart JM (2013) Cancer panomics: computational methods and infrastructure for integrative analysis of cancer high-throughput "omics" data-session introduction. In: Biocomputing 2014. World Scientific, Singapore, pp 1–2
72. Melloni GE, Ogier AG, de Pretis S et al (2014) DOTS-Finder: a comprehensive tool for assessing driver genes in cancer genomes. Genome Med 6:44. https://doi.org/10.1186/gm563
73. Wong K, Keane TM, Stalker J, Adams DJ (2010) Enhanced structural variant and breakpoint detection using SVMerge by integration of multiple detection methods and local assembly. Genome Biol 11:R128. https://doi.org/10.1186/gb-2010-11-12-r128
74. Akavia UD, Litvin O, Kim J et al (2010) An integrated approach to uncover drivers of cancer. Cell 143:1005–1017. https://doi.org/10.1016/j.cell.2010.11.013
75. Sanchez-Garcia F, Villagrasa P, Matsui J et al (2014) Integration of genomic data enables selective discovery of breast cancer drivers. Cell 159:1461–1475. https://doi.org/10.1016/j.cell.2014.10.048
76. Chen Y, Hao J, Jiang W et al (2013) Identifying potential cancer driver genes by genomic data integration. Sci Rep 3:3538. https://doi.org/10.1038/srep03538
77. Schroeder MP, Rubio-Perez C, Tamborero D et al (2014) OncodriveROLE classifies cancer driver genes in loss of function and activating mode of action. Bioinformatics 30:i549–i555. https://doi.org/10.1093/bioinformatics/btu467
78. Bertrand D, Chng KR, Sherbaf FG et al (2015) Patient-specific driver gene prediction and risk assessment through integrated network analysis of cancer omics profiles. Nucleic Acids Res 43:e44. https://doi.org/10.1093/nar/gku1393
79. Berman HM, Westbrook J, Feng Z et al (2000) The protein data bank. Nucleic Acids Res 28:235–242. https://doi.org/10.1093/nar/28.1.235
80. Jia P, Wang Q, Chen Q et al (2014) MSEA: detection and quantification of mutation hotspots through mutation set enrichment analysis. Genome Biol 15:489. https://doi.org/10.1186/s13059-014-0489-9
81. Cheng F, Jia P, Wang Q, Zhao Z (2014) Quantitative network mapping of the human kinome interactome reveals new clues for rational kinase inhibitor discovery and individualized cancer therapy. Oncotarget 5:3697–3710. https://doi.org/10.18632/oncotarget.1984
82. Reimand J, Bader GD (2013) Systematic analysis of somatic mutations in phosphorylation signaling predicts novel cancer drivers. Mol Syst Biol 9:637. https://doi.org/10.1038/msb.2012.68
83. Wang Y, Cheng H, Pan Z et al (2015) Reconfiguring phosphorylation signaling by genetic polymorphisms affects cancer susceptibility. J Mol Cell Biol 7:187–202. https://doi.org/10.1093/jmcb/mjv013
84. Vuong H, Cheng F, Lin C-C, Zhao Z (2014) Functional consequences of somatic mutations in cancer using protein pocket-based prioritization approach. Genome Med 6:81. https://doi.org/10.1186/s13073-014-0081-7
85. Zhao J, Cheng F, Wang Y et al (2016) Systematic prioritization of druggable mutations in ~5000 genomes across 16 cancer types using a structural genomics-based approach. Mol Cell Proteomics 15:642–656. https://doi.org/10.1074/mcp.M115.053199
86. Ghersi D, Singh M (2014) Interaction-based discovery of functionally important genes in cancers. Nucleic Acids Res 42:e18. https://doi.org/10.1093/nar/gkt1305
87. Ryslik GA, Cheng Y, Cheung K-H et al (2013) Utilizing protein structure to identify non-random somatic mutations. BMC Bioinf 14:190. https://doi.org/10.1186/1471-2105-14-190

88. Porta-Pardo E, Godzik A (2014) e-Driver: a novel method to identify protein regions driving cancer. Bioinformatics 30:3109–3114. https://doi.org/10.1093/bioinformatics/btu499
89. Pe'er D, Hacohen N (2011) Principles and strategies for developing network models in cancer. Cell 144:864–873. https://doi.org/10.1016/j.cell.2011.03.001
90. Ng S, Collisson EA, Sokolov A et al (2012) Paradigm-shift predicts the function of mutations in multiple cancers using pathway impact analysis. Bioinformatics 28:i640–i646. https://doi.org/10.1093/bioinformatics/bts402
91. Paull EO, Carlin DE, Niepel M et al (2013) Discovering causal pathways linking genomic events to transcriptional states using tied diffusion through interacting events (TieDIE). Bioinformatics 29:2757–2764. https://doi.org/10.1093/bioinformatics/btt471
92. Bashashati A, Haffari G, Ding J et al (2012) DriverNet: uncovering the impact of somatic driver mutations on transcriptional networks in cancer. Genome Biol 13:R124. https://doi.org/10.1186/gb-2012-13-12-r124
93. Hou JP, Ma J (2014) DawnRank: discovering personalized driver genes in cancer. Genome Med 6:56. https://doi.org/10.1186/s13073-014-0056-8
94. Jia P, Zhao Z (2014) VarWalker: personalized mutation network analysis of putative cancer genes from next-generation sequencing data. PLoS Comput Biol 10:e1003460. https://doi.org/10.1371/journal.pcbi.1003460
95. Hofree M, Shen JP, Carter H et al (2013) Network-based stratification of tumor mutations. Nat Methods 10:1108–1115. https://doi.org/10.1038/nmeth.2651
96. Vandin F, Upfal E, Raphael BJ (2011) Algorithms for detecting significantly mutated pathways in cancer. J Comput Biol 18:507–522. https://doi.org/10.1089/cmb.2010.0265
97. Stumpf MPH, Thorne T, de Silva E et al (2008) Estimating the size of the human interactome. Proc Natl Acad Sci U S A 105:6959–6964. https://doi.org/10.1073/pnas.0708078105
98. Menche J, Sharma A, Kitsak M et al (2015) Disease networks. Uncovering disease-disease relationships through the incomplete interactome. Science 347:1257601. https://doi.org/10.1126/science.1257601
99. Perco P, Rapberger R, Siehs C et al (2006) Transforming omics data into context: bioinformatics on genomics and proteomics raw data. Electrophoresis 27:2659–2675. https://doi.org/10.1002/elps.200600064
100. Cavin Périer R, Junier T, Bucher P (1998) The eukaryotic promoter database EPD. Nucleic Acids Res 26:353–357. https://doi.org/10.1093/nar/26.1.353
101. Martens L, Hermjakob H, Jones P et al (2005) PRIDE: the proteomics identifications database. Proteomics 5:3537–3545. https://doi.org/10.1002/pmic.200401303
102. Flicek P, Aken BL, Beal K et al (2008) Ensemble 2008. Nucleic Acids Res 36:D707–D714. https://doi.org/10.1093/nar/gkm988
103. Karolchik D, Kuhn RM, Baertsch R et al (2008) The UCSC genome browser database: 2008 update. Nucleic Acids Res 36:D773–D779. https://doi.org/10.1093/nar/gkm966
104. Maglott D, Ostell J, Pruitt KD, Tatusova T (2005) Entrez gene: gene-centered information at NCBI. Nucleic Acids Res 33:D54–D58. https://doi.org/10.1093/nar/gki031
105. Camon E, Magrane M, Barrell D et al (2004) The Gene Ontology Annotation (GOA) database: sharing knowledge in Uniprot with Gene Ontology. Nucleic Acids Res 32:D262–D266. https://doi.org/10.1093/nar/gkh021
106. Apweiler R, Bairoch A, Wu CH et al (2004) UniProt: the universal protein knowledgebase. Nucleic Acids Res 32:D115–D119. https://doi.org/10.1093/nar/gkh131
107. Sterk P, Kersey PJ, Apweiler R (2006) Genome reviews: standardizing content and representation of information about complete genomes. OMICS 10:114–118. https://doi.org/10.1089/omi.2006.10.114
108. Ogata H, Goto S, Sato K et al (1999) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 27:29–34. https://doi.org/10.1093/nar/28.1.27
109. Ashburner M, Ball CA, Blake JA et al (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25–29. https://doi.org/10.1038/75556

110. DeJongh M, Formsma K, Boillot P et al (2007) Toward the automated generation of genome-scale metabolic networks in the SEED. BMC Bioinf 8:139. https://doi.org/10.1186/1471-2105-8-139
111. Krieger CJ, Zhang P, Mueller LA et al (2004) MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nucleic Acids Res 32:D438–D442. https://doi.org/10.1093/nar/gkh100
112. Karp PD, Ouzounis CA, Moore-Kochlacs C et al (2005) Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. Nucleic Acids Res 33:6083–6089. https://doi.org/10.1093/nar/gki892
113. Ren Q, Chen K, Paulsen IT (2007) TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. Nucleic Acids Res 35:D274–D279. https://doi.org/10.1093/nar/gkl925
114. Zhao F, Xuan Z, Liu L, Zhang MQ (2005) TRED: a transcriptional regulatory element database and a platform for in silico gene regulation studies. Nucleic Acids Res 33:D103–D107. https://doi.org/10.1093/nar/gki004
115. Ji ZL, Chen X, Zhen CJ et al (2003) KDBI: kinetic data of bio-molecular interactions database. Nucleic Acids Res 31:255–257
116. Le Novère N, Bornstein B, Broicher A et al (2006) BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. Nucleic Acids Res 34:D689–D691. https://doi.org/10.1093/nar/gkj092
117. Sivakumaran S, Hariharaputran S, Mishra J, Bhalla US (2003) The database of quantitative cellular signaling: management and analysis of chemical kinetic models of signaling networks. Bioinformatics 19:408–415. https://doi.org/10.1093/bioinformatics/btf860
118. Xenarios I, Salwínski L, Duan XJ et al (2002) DIP, the database of interacting proteins: a research tool for studying cellular networks of protein interactions. Nucleic Acids Res 30:303–305
119. Zanzoni A, Montecchi-Palazzi L, Quondam M et al (2002) MINT: a molecular interaction database. FEBS Lett 513:135–140
120. Pagel P, Kovac S, Oesterheld M et al (2005) The MIPS mammalian protein-protein interaction database. Bioinformatics 21:832–834. https://doi.org/10.1093/bioinformatics/bti115
121. Edgar R, Domrachev M, Lash AE (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30:207–210. https://doi.org/10.1093/nar/30.1.207
122. Sherlock G, Hernandez-Boussard T, Kasarskis A et al (2001) The stanford microarray database. Nucleic Acids Res 29:152–155
123. Shannon P, Markiel A, Ozier O et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/gr.1239303
124. Shannon PT, Reiss DJ, Bonneau R, Baliga NS (2006) The Gaggle: an open-source software system for integrating bioinformatics software and data sources. BMC Bioinf 7:176. https://doi.org/10.1186/1471-2105-7-176
125. Wolff AC, Hammond MEH, Hicks DG et al (2013) Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. J Clin Oncol 31:3997–4013. https://doi.org/10.1200/JCO.2013.50.9984
126. van de Vijver MJ, He YD, van't Veer LJ et al (2002) A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347:1999–2009. https://doi.org/10.1056/NEJMoa021967
127. Paik S, Shak S, Tang G et al (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 351:2817–2826. https://doi.org/10.1056/NEJMoa041588
128. Sorlie T, Tibshirani R, Parker J et al (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100:8418–8423. https://doi.org/10.1073/pnas.0932692100

129. Arnedos M, Vicier C, Loi S et al (2015) Precision medicine for metastatic breast cancer--limitations and solutions. Nat Rev Clin Oncol 12:693–704. https://doi.org/10.1038/nrclinonc.2015.123
130. Ligthart S, Vaez A, Võsa U et al (2018) Genome analyses of >200,000 individuals identify 58 loci for chronic inflammation and highlight pathways that link inflammation and complex disorders. Am J Hum Genet 103:691–706. https://doi.org/10.1016/j.ajhg.2018.09.009
131. Abrahams E (2008) Right drug-right patient-right time: personalized medicine coalition. Clin Transl Sci 1:11–12. https://doi.org/10.1111/j.1752-8062.2008.00003.x
132. Metias SM, Lianidou E, Yousef GM (2009) MicroRNAs in clinical oncology: at the crossroads between promises and problems. J Clin Pathol 62:771–776. https://doi.org/10.1136/jcp.2009.064717
133. Ashley EA, Butte AJ, Wheeler MT et al (2010) Clinical assessment incorporating a personal genome. Lancet 375:1525–1535. https://doi.org/10.1016/S0140-6736(10)60452-7
134. Abelson S, Collord G, Ng SWK et al (2018) Prediction of acute myeloid leukaemia risk in healthy individuals. Nature 559:400–404. https://doi.org/10.1038/s41586-018-0317-6
135. Klein RJ, Zeiss C, Chew EY et al (2005) Complement factor H polymorphism in age-related macular degeneration. Science 308:385–389. https://doi.org/10.1126/science.1109557
136. Johnson AD, O'Donnell CJ (2009) An open access database of genome-wide association results. BMC Med Genet 10:6. https://doi.org/10.1186/1471-2350-10-6
137. Cancer Genome Atlas Network (2012) Comprehensive molecular characterization of human colon and rectal cancer. Nature 487:330–337. https://doi.org/10.1038/nature11252
138. Stephens PJ, Tarpey PS, Davies H et al (2012) The landscape of cancer genes and mutational processes in breast cancer. Nature 486:400–404. https://doi.org/10.1038/nature11017
139. Natter MD, Quan J, Ortiz DM et al (2013) An i2b2-based, generalizable, open source, self-scaling chronic disease registry. J Am Med Inform Assoc 20:172–179. https://doi.org/10.1136/amiajnl-2012-001042
140. Stenson PD, Ball EV, Howells K et al (2009) The human gene mutation database: providing a comprehensive central mutation database for molecular diagnostics and personalised genomics. Hum Genomics 4:69. https://doi.org/10.1186/1479-7364-4-2-69
141. Chen R, Mias GI, Li-Pook-Than J et al (2012) Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell 148:1293–1307. https://doi.org/10.1016/j.cell.2012.02.009
142. Pasic MD, Samaan S, Yousef GM (2013) Genomic medicine: new frontiers and new challenges. Clin Chem 59:158–167. https://doi.org/10.1373/clinchem.2012.184622
143. Fischer HP (2008) Mathematical modeling of complex biological systems: from parts lists to understanding systems behavior. Alcohol Res Health 31:49–59
144. Vera J, Wolkenhauer O (2008) A system biology approach to understand functional activity of cell communication systems. Methods Cell Biol 90:399–415