Molecular & Cellular Oncology 1:2, e957981; April/May/June 2014; Published with license by Taylor & Francis Group, LLC

Approaches to uncovering cancer diagnostic and prognostic molecular signatures

Shengjun Hong†, Yi Huang†, Yaqiang Cao†, Xingwei Chen†, and Jing-Dong J Han*

Chinese Academy of Sciences Key Laboratory of Computational Biology; Chinese Academy of Sciences-Max Planck Partner Institute for Computational Biology; Shanghai Institutes for Biological Sciences; Chinese Academy of Sciences; Shanghai, China

†These authors contributed equally to this work.

Keywords: Bayesian network, cancer diagnosis, cancer prognosis, molecular signature, subtype

Abbreviations: BIC, Bayesian information criterion; BN, Bayesian network; CNV, copy number variation; COPA, cancer outlier profile analysis; DEG, differentially expressed genes; DMV, DNA methylation valley; GSEA, gene set enrichment analysis; GO, gene ontology; GWAS, genome-wide association studies; lncRNA, long non-coding RNA; PAGE, parametric analysis of gene set enrichment; PheWAS, phenome-wide association studies; SOM, self-organizing map; SNP, single-nucleotide polymorphism; SVM-RFE, Support Vector Machine-Recursive Feature Elimination; TCGA, The Cancer Genome Atlas; TCIA, The Cancer Imaging Archive.

The recent rapid development of high-throughput technology enables the study of molecular signatures for cancer diagnosis and prognosis at multiple levels, from genomic and epigenomic to transcriptomic. These unbiased large-scale scans provide important insights into the detection of cancer-related signatures. In addition to single-layer signatures, such as gene expression and somatic mutations, integrating data from multiple heterogeneous platforms using a systematic approach has been proven to be particularly effective for the identification of classification markers. This approach not only helps to uncover essential driver genes and pathways in the cancer network that are responsible for the mechanisms of cancer development, but will also lead us closer to the ultimate goal of personalized cancer therapy.

Introduction

Cancer is a major human health problem worldwide and is related to one-fourth of deaths in the United States.[1] Decades of cancer research have revealed many specific details, as well as some general features shared among different cancers. Although the rate of cancer deaths in the United States has declined in recent decades,[1,2] the rate of acquisition of cancer is continuously increasing.[1,3] Early detection of cancer diagnosis signatures decreases both morbidity and mortality. Moreover, studying the signatures associated with cancer prognosis (quantified by 5-year survival time in clinical trials) not only helps to predict patient outcome, but also holds a key to understanding the genetic mechanisms of cancer development.[4,5]

With the development of high-throughput technology, signatures existing at multiple levels have been identified for cancer diagnosis and prognosis, including genomic, epigenomic, and transcriptomic signatures. For example, genome-wide single nucleotide polymorphism (SNP) profiling and array-based comparative genomic hybridization have been applied to identify germline and somatic lesions in several cancers. Additionally, hundreds of SNPs or haplotypes have been reported to be significantly associated with cancers. In a study of 1,599 cases and 11,546 controls, Stacey et al.[6] found that rs3803662, which is associated with the *TOX3* gene, is also significantly associated with breast cancer. The DNA methylome is also under intense study, providing global pictures of epigenetic changes in cancers.[7] Transcriptome analyses have successfully demonstrated that the expression of multiple genes, rather than single genes, can serve as effective subtype or prognosis classifiers for many cancers, such as leukemia[8] and breast cancer.[9]

Although different layers of genome-wide analysis have revealed global features of cancers, integration of multilayer information facilitates more accurate cancer subtyping and more comprehensive mechanistic insights. Within such a panorama, the systematic approach has led to identification of the hallmarks of cancers.[10]

In this review, we aim to provide an insight into the data and methods available for systems level analysis of cancer subtypes and their characterization. We have organized the

*Correspondence to: Jing-Dong J Han; Email: jdhan@picb.ac.cn

Submitted: 06/06/2014; Revised: 07/21/2014; Accepted: 07/22/2014
http://dx.doi.org/10.4161/23723548.2014.957981

review into 3 parts: first, we introduce general features and web-based resources for molecular signatures of cancer diagnosis and prognosis; second, we summarize existing methods for detecting such signatures; and, finally, we discuss potential methods for interpreting these signatures, such as network and module analysis.

Cancer-related high-throughput data types and web resources

Although it is feasible to collect raw cancer-related high-throughput data, such as from GEO[11] and ArrayExpress,[12] several databases and web services provide rich cancer-related data in a curated or integrated manner.

The Cancer Genome Atlas (TCGA) project has generated a myriad of cancer "omic" data. To date, more than 8,913 tumor samples across 30 types of cancer have been collected and sequenced. The TCGA provides raw and processed data covering layers of genome, epigenome, and transcriptome data, together with clinical information. The recently established cBioPortal[13] provides not only downloadable large-scale cancer genomic data, but also online visualization and analysis services for TCGA datasets.

In addition to these comprehensive resources, there are several databases focusing on 1 or 2 specific areas. For example, COSMIC[14] stores somatic mutations. Its latest version presents a cancer mutation landscape of 132 known cancer genes and 208 fusion gene pairs, based on nearly 8,000 cancer genomes. With a convenient interface, COSMICMart[15] helps to filter COSMIC data sets into categories. Oncomine[16] is a database for target identification and validation, drug development, and clinical research. Oncotator (http://www.broadinstitute.org/oncotator/) provides annotation for cancer genes, mutations, and amplification or deletion regions. Tumorscape[17] provides both a portal to query copy number alterations across multiple cancer types, and a web interface visualizing the results based on the GISTIC[18] algorithm. IntOgen[19] integrates somatic mutations, copy number changes, and expression in cancer into 3 query-and-download modules, in addition to providing an interface to TCGA.

These valuable resources have facilitated various efforts in cancer studies and have broadened our perspectives of cancer (**Table 1**). At the present time most resources are based on isolated samples but we expect that there will be an increase in replicated data for advanced analyses, such as multiple (and even time series) samples from the same patient or from a homogenous population.

Over the past few years, multiple types of signatures have been reported to associate with cancer diagnosis and prognosis. Most of these studies focused on common somatic mutations,[20] mRNA[21,22] and microRNA expression,[23,24] and protein level changes.[25,26] With the popularization of high-throughput sequencing technologies and the refinement of bioinformatics pipelines, these features have helped to identify credible markers.

In addition to general genomic and transcriptomic features, various other features could be used to improve the confidence. For example, long non-coding RNA (lncRNA) is an emerging new paradigm in cancer research that can be either oncogenic or tumor suppressive, indicating its possible application for diagnostics and prognosis. One typical example for practical diagnostics is PCA3,[27] which is widely used in urine testing to determine prostate cancer risk. HOTAIR, which has a chromatin-remodeling effect, serves as an oncogenic biomarker and has been validated in a variety of cancers, such as lung cancer[28] and liver cancer.[29] MEG3 acts as a tumor suppressor that is frequently downregulated in pituitary cancer[30] and glioma.[31] lncRNAs have advantages over protein-coding RNAs in cancer diagnosis and prognosis because of their expression specificity and direct molecular function.

The methylome, or DNA methylation status on genome-wide CpG sites, has been intensively studied in developmental biology. However, despite the fact that cancer shares key properties with development, albeit inversely, the methylome was only recently applied to cancer diagnosis or prognosis. It has been reported that GSTP1 gains hypermethylation in prostate cancer, indicating its role as a diagnostic marker.[32] A recent study showed that DNA methylation valleys (DMVs) in stem cells are hypermethylated in cancer[33] and therefore provide novel aberrant signatures. Nearly 8,000 cancer methylomes are available through public databases,[34] facilitating future studies to reveal the specific DNA methylation signatures that drive carcinogenesis.

Clinical features are useful tools for diagnosis and prognosis. In particular, imaging has been widely applied in cancer diagnosis using various systems[35] such as X-ray, computed tomography scan, magnetic resonance imaging, tumor biopsy, and endoscopic examination. The Cancer Imaging Archive (TCIA)[36] contains medical images of both the National Lung Screening Trial

Table 1. Summary of cancer data resources

Web Resource	Raw data	Preprocessed data	Features	Clinical information
TCGA and cBioPortal	Yes	Yes	Genomic, transcriptomic, epigenomic, proteomic	Yes
COSMIC & COSMICMart	No	Yes	Mutation	No
Oncomine	Yes	Yes	Microarray-based gene expression and copy number variation	Yes
Oncotator	No	Yes	Gene, mutation, cancer amplification, and deletion region	No
Tumorscape	No	Yes	Copy number variation	No
TCIA	Yes	Yes	Medical images	Yes

project[37] and the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial project.[38] These clinical features reflect obvious biological outcomes and should also assist molecular signature identification.

Approaches

As resources and signature types of cancers keep emerging, so do various bioinformatics methods for analyzing such data. We will introduce the approaches that are most frequently used for either associative or mechanistic inferences.

Associative inference

Cancer diagnosis and prognosis have benefitted from the application of multiple molecular markers. However, given the heterogeneous nature of cancers, prognosis-related subtype classification is more, or at least equally, important for personalized treatment.[39] Methods for subtype detection and classification can be generally grouped into supervised and unsupervised approaches, by which subtypes are characterized by certain signatures (**Fig. 1**). When differentiating clinically well-defined subtypes, supervised approaches are often used. However, for identification of unknown subtypes or when classifying clinically less well-defined subtypes, unsupervised methods are required.[40]

Supervised learning of molecular signatures

Supervised learning is quite straightforward when the samples are well categorized, for example between cancer samples and controls. In this approach, known subtype-related signals are extracted from noise or confounding factors.

Genetic variation in the human genome provides an abundant resource for cancer research. The study of genetic variation can reveal susceptibility to cancer associated with distinct variations.[41] Genome-wide association studies (GWAS) have identified many genetic risk factors for different common human cancers, which are available in the NHGRI GWAS catalog.[42] The most frequently used statistics in GWAS are the Chi-square test and Fisher's exact test, although the Chi-square test is not suitable when the sample size is small (fewer than 10 samples). As the Pearson Chi square test cannot handle a variable with more than 2 categories, the Cochran-Armitage test for trend may be applied in such cases. All of these statistics tests are integrated in PLINK,[43] a popular tool for GWAS. Not all SNPs can be accurately genotyped, and SNPTEST[44] uses Frequentist and Bayesian tests to solve this problem. As GWAS datasets generally involve more than 100,000 SNPs, multiple testing correction is needed. The commonly used correction method is the Bonferroni correction but this may be too strong in some cases, for which the Holm–Bonferroni method might be more suitable. There is no generic bridge that links SNPs to cancers; only when integrated with functional data can a cancer-related SNP be deemed to be responsible for the cancer process.[45] Phenome-wide association studies (PheWAS)[46] can be viewed as a variant of GWAS to investigate the associations between SNPs and phenotypes. This method, especially when accompanied by imaging, is a promising approach to explore cancer genome–phenome associations.

Unlike genomic data, the most general approach to transcriptome data is to identify differentially expressed genes (DEGs). The Student's t test is frequently used for analysis of DEGs. The t statistic assumes homogeneity of examined samples; however, this does not always hold true in cancer samples. For example, an unstable genome may lead to the same translocations (and subsequent abnormal expression) in some, but not all, cancer samples, and not in the control samples.[47] To solve this problem, methods that are sensitive to "outliers" have been developed. Cancer outlier profile analysis (COPA) was proposed and applied to prostate cancer,[47,48] and then implemented as a part of the abovementioned Oncomine. In addition to COPA, several statistics have been sequentially proposed, including OS,[49] ORT,[50] MOST,[51] and GTI.[52] These statistics may work differently depending on the type of data and thus should be carefully compared.[53]

Alternatively, sometimes the data may itself contain homogeneity (e.g., after proper subtyping), or researchers may be

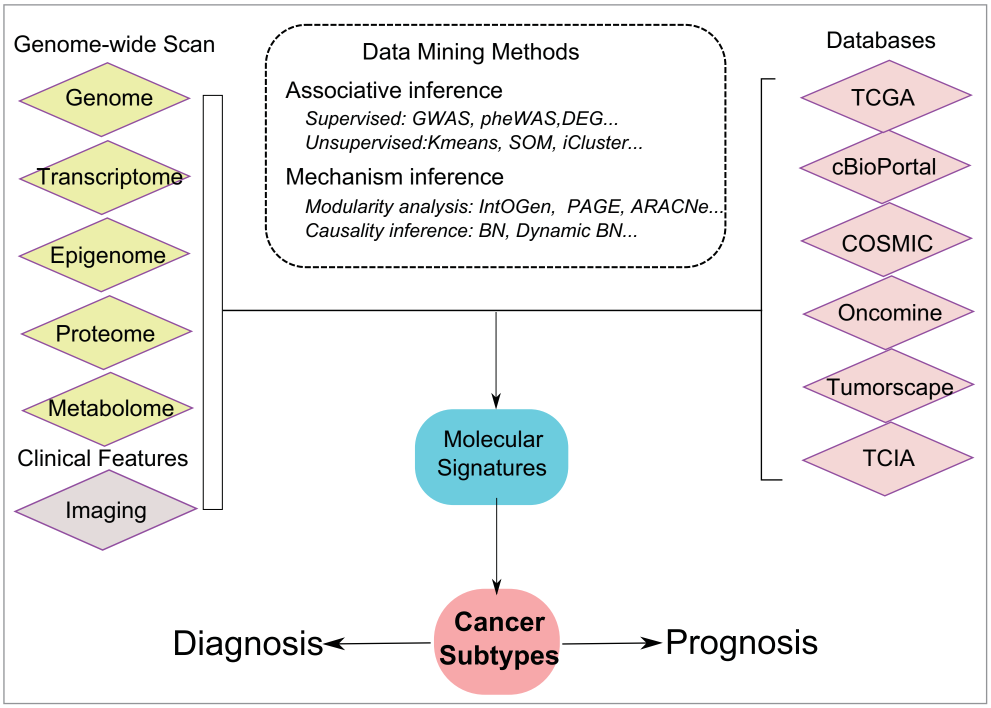

Figure 1. Overview of strategies to detect cancer subtypes related to cancer diagnosis and prognosis.

interested in only the signatures with high penetrance. In these cases, in addition to the Student's t test there are many well-developed tools that could be applied, such as SAM,[54] limma,[55] edgeR,[56] DESeq,[57] and Cuffdiff.[58] If researchers are unsure about the homogeneity, machine-learning approaches may be more robust. For example, we have applied Support Vector Machine-Recursive Feature Elimination (SVM-RFE) to identify markers of pediatric acute lymphoblastic leukemia.[59]

When data come from multiple resources, either generated in different batches or compared across multiple layers, it is essential to properly preprocess the pooled data and extract signals from noise and confounding factors using so-called "meta-analysis".[60,61] Based on the extent of noise reduction, preprocessing methods can be divided into 3 categories, as follows: (1) Moderate approaches smooth data with different relative intensities either by z-score normalization or quantile normalization. Z-score normalization requires the original data to have an approximately Gaussian distribution whereas quantile normalization is non-parametric and especially efficient in microarray data analysis (implemented in RMA).[62] Although both work on sample noise reduction, z-score normalization could also be applied to features, transforming expression intensities to expression patterns.[59] This is often a necessary preprocessing step to integrate time-series or multilayer data.[59,63] (2) Known confounding factors, typically batch effects, can be reduced after being incorporated into the null model using a generalized linear model such as in DESeq,[57] or an empirical Bayes method such as in Combat.[64] (3) It is also possible to further exclude unknown confounding factors, as implemented in SVA[65] and ISVA.[66]

Unsupervised learning of molecular signatures

Unsupervised methods for subtyping may be more intrinsic and more robust for experimental designs. Clustering is widely used for omics data, to divide features or samples into subgroups. Since the first application of hierarchical clustering in microarray gene expression datasets,[67] the approach has rapidly been applied to cancer gene expression datasets.[8,68] K-means clustering and Self-Organizing Maps (SOMs) are often applied in similar ways. At the present time hierarchical clustering remains the typical choice; however, it is challenged by multilayer data. Efforts have therefore been made to improve clustering for integrative analysis, for example classic hierarchical clustering of the correlation matrix between mRNA and microRNA expression[69] and biclustering on the correlation matrix between mRNA expression and DNA copy number.[70] More sophisticated tools were also developed, such as iCluster[71] and its extended version iClusterPlus,[72] PSDF,[73] MDI,[74] JIVE,[75] and SNF.[76] iCluster is intensely used with TCGA[77,78] to discover subtypes with distinct clinical outcomes, whereas iClusterPlus is able to handle both discrete (somatic mutations) and continuous variables. Most cluster methods cannot automatically determine the optimal number of sample and feature clusters. We proposed an adaptive clustering algorithm incorporating the Bayesian information criterion (BIC) and an unsupervised "Super k-means" method. With this approach, we detected 7 subtypes for ovarian cancer with significantly different clinical outcomes based on the combination of mRNA and miRNA expression, DNA methylation, and copy number variation. Accordingly, we also developed a useful framework to detect modular signatures for prognosis.[79]

Mechanism inference

As mentioned above, there are various features and resources that provide multiple dimensions of signatures for cancer diagnosis and prognosis. However, association only provides a pool of molecules, including false positives, unimportant candidates that should be ignored, and false negatives. Moreover, cancer arises from a complex origin composed of genomic, transcriptomic, and epigenomic variations.[80] Disease-specific genes do not function independently; instead, they usually act as a network module that associates with a certain biological function.[81,82] In several cases, a single type of molecular signature cannot uncover the molecular mechanism of cancer prognosis to predict clinical outcome.[83] In order to either accurately diagnose disease at an early stage or pursue the cause of prognosis, careful refinement of these candidate signatures is required, especially approaches based on networks with modularity analysis or causality inference.[84]

Modularity analysis

Modularity is an important feature of networks. A network module is a context-coherent sub-network with conditionally comparable temporal and spatial profiles, and ideally with defined inputs and outputs.[85] In practice, modules are often described without the exact inputs and outputs defined (but instead assumed) and evaluated by relative functional homogeneity within the module.[82] Based on the static structure of a network, a module can be defined and dissected as a sub-network whose nodes are more densely connected within the sub-network than toward the outside, or that has more than random expectation as measured by the modularity metric[86] or the clustering coefficient.[87]

The cancer signaling network, for example, can be topologically divided into 12 blocks or modules.[88] Edge type consistency has been used to find epistatic modules,[89] although gene expression profile similarity and dissimilarity are most frequently used to define dynamic modules that are active under a certain biological context.[90]

With no exception, classification markers are modularly organized and reflect dysregulation or driver mutations through network analysis. An unbiased search for markers of pediatric acute lymphoblastic leukemia subtype classification yielded a group of 62 genes that segregate into several network modules for different subtypes and are potentially controlled by subtype-specific transcriptional regulators.[59] Such modularity in cancer marker genes has been used to identify network-based classifiers that have been demonstrated to outperform classical single gene or non-network module-based classifiers. Compared with gene expression networks, markers identified as sub-networks from protein-protein interaction networks are more reproducible and achieve significantly higher accuracy for classifying metastatic versus non-metastatic breast tumors.[91] Martin et al. have found that highquality breast cancer prognosis markers can only be identified within subtypes, and that combinations of various markers can optimize the performance of the marker gene set. Most surprisingly, they found that each marker gene signature forms a network module, within which the marker genes interact intensively with genes that are frequently mutated in breast cancers, although the marker genes themselves are mostly not mutated. Moreover, the mutated interacting genes in the modules can also distinguish metastatic *versus* non-metastatic samples, implying that these might be driver mutations within each module.[92] Therefore, the modularity of disease-associated genes in molecular interaction networks allows prediction of new disease-associated genes through their direct or indirect interactions with known disease-associated genes.

Uncovering the cancer regulatory network without a predefined reference set will reduce study bias and facilitate more objective analysis of all potential features involved in the network properties. To this end, Andreas Califano's group has developed an algorithm called "ARACNe" which can *ab initio* infer regulatory interactions based on mutual information between 2 genes across a set of measurements.[93] In particular, they have successfully applied the algorithm to predict transcriptional interactions specific for high-grade glioma.[93] In combination with searches for transcription factors whose targets overlap significantly with genes that are overexpressed in mesenchymal cells, they narrowed down a key regulatory module.[94] Also using an approach based on mutual information, Mani et al. developed a new algorithm that can identify dysregulated interactions in B-cell lymphoma using a Bayesian analysis that predicted a B-cell specific interactome as the backbone. The dysregulation is defined as loss/gain of a correlation in gene expression comparing lymphoma versus normal B cells.[95] Compared with candidate gene-based reverse engineering approaches, such *de novo* network reverse engineering can identify not only dysregulated interactions, but also coherently dysregulated network modules that arise from the interactions, in an unbiased manner.[95]

Most of the abovementioned methods of network modularity analysis have been implemented in Cytoscape[96] and its rich plugin apps.[97] Cytoscape is an expanding computational platform for the integration, visualization, statistical modeling, and annotation of biological networks.[98] The apps are well organized and categorized within http://apps.cytoscape.org/, which makes its convenient for users to access and use to compile an analysis pipeline.

After modularity inference the general network modules are always compared to specific biology "pathways," such as widely used Gene Ontology (GO) terms[99] and KEGG pathways.[100] This comparison is termed enrichment analysis, and often uses Fisher's exact test or hypergeometric test to draw statistical significance for the selected enriched terms. For GO analysis only, AmiGO[101] and BiNGO[102] implemented in Cytoscape are good choices. DAVID[103] is highly recommended for multisource integration. The IntOGen web application allows evaluation of the contribution of biological modules such as KEGG pathways to a cancer by testing the significance of overlap between genes that are changed in the cancer and genes in a defined module.[104] More general-purpose applications such as Gene set enrichment analysis (GSEA)[105] or its modified version Parametric analysis of gene set enrichment (PAGE)[106] can also reveal whether a pathway, module, or signature set is significantly changed based on the rank or average expression intensity of genes within a gene set.

Bayesian networks and causality inference

Compared with mutual information or correlation-based methods, Bayesian network (BN) inference as a network reverse engineering approach has higher theoretical consistency, is able to distinguish direct and indirect interactions, and can identify both strong and weak and linear and non-linear dependencies as well as potential causal relationships.[107,108] BN is a network or graph representation of the joint probability distribution over a set of variables (nodes) or conditional dependencies between variables. The BN structural learning algorithm searches for the network structure that has the best fit of joint probability distribution to the data using a scoring function such as the BIC. BIC contains 2 terms: one to evaluate the likelihood that the data are generated by the model, and another to penalize the complexity of the model.[109] Recently, BN has been used for the diagnosis and prognosis of several cancer types,[110] including breast cancer[111] and lung cancer.[112] Olivier et al. applied BN to integrate clinical and microarray data.[111] Evaluation of the performance of BN showed that this method performed well in predicting prognosis of breast cancer patients.

One restriction of BN learning is that the graph must be acyclic; that is, no loops are allowed even though they truly exist. Such feedback relationships can sometimes be resolved by additional temporal information, for example by the so-called "dynamic BN" approach. Potential causal relationships can also be identified from the consistently directed edges (irreversible edges) within the whole set of equivalent BN structures.[113] Data from gene perturbation experiments can provide more direct evidence for inferring causal relationships. For example, a directed signaling network of 11 molecules can be reverse engineered by BN learning on thousands of single-cell flow cytometry measurements of the level of the molecules in human primary T cells after gene perturbations in the network.[108]

The requirement for a large number of data points for BN inference has been a limiting factor in directly inferring gene regulatory networks from gene expression measurements. The recent rapid accumulation of microarray and deep sequencing data has made such approaches more practical. In pursuing key signatures, "early" changes may arise from system instability and thus have low penetrance, whereas a few function-related "late" changes are causal to cancer development. Therefore, the "intermediate" mediators and potential causality inferred by BNs will facilitate both early diagnoses and accurate prognoses when the core of the network is found to be affected by

Table 2. Methods for the detection of molecular signatures of cancer diagnosis and prognosis

	Approach	Summary	Data type
Associative inference–Supervised learning			
GWAS	PLINK	An open-source whole genome association tool, including statistics such as Chi-square test, Cochran-Armitage test, and Fisher's exact test.	Genotype, CNV, and haplotype
	SNPTEST	Incorporates imputation methods for genotype association test	Genotype
PheWAS		Investigates the association between SNP and phenotypes	Genotype and phenotype
DEGs	Student's t-test, SAM, limma, edgeR, DESeq, Cuffdiff	Identifies DEGs, assuming homogeneity of examined samples	Gene expression
	COPA, OS, ORT, MOST, GTI, SVM-RFE	Identifies DEGs, robust to heterogeneity of examined samples	
Noise reduction	Z-score normalization	Preprocess expression data with relative intensities	Multiple-layer data integration
	Quantile normalization		
	Combat	Handles known confounding factors such as batch effects	Single-layer data
	SVA and ISVA	Excludes unknown confounding factors	Single-layer data
Associative inference–Unsupervised learning			
Clustering analysis	Hierarchical clustering, Kmeans clustering, SOM	Partitions features or samples into subgroups	Single-layer data
	Biclustering, iCluster, iClusterPlus, PSDF, MDI, JIVE, SNF, Super k-means	Discovers subtypes with clinical outcomes, integrating multiple types of data	Multiple-layer data integration
Mechanism inference–Modularity analysis			
Subnetwork function analysis	IntOGen	Evaluates the contribution of biological modules to a cancer	Gene-gene association
	DAVID, GSEA, PAGE	Reveals whether a cancer-related module is significantly enriched for a known pathway	Gene expression
	ARACNe	Infers regulatory interactions based on mutual information between genes	Gene expression
	Cytoscape	Network and modularity analysis	Multiple-layer data integration
Mechanism inference—Bayesian network and causality inference			
De novo network	Bayesian network	Infers a network by detecting potential causal relationships between genes	Multiple-layer data integration
	Dynamic BN	Allows feedback relationship compared to regular Bayesian network	Multiple-layer data integration

Abbreviations: BN, Bayesian network; CNV, copy number variation; COPA, cancer outlier profile analysis; DEG, differentially expressed genes; GSEA, gene set enrichment analysis; PAGE, parametric analysis of gene set enrichment.

perturbation.[110] The approaches discussed here are summarized in **Table 2**.

Discussion and Perspectives

With the huge amount of high-throughput data that is already available or is being generated at an accelerated rate for different layers of the cancer molecular interaction network, obtaining a global picture of the full cancer molecular network for each cancer type, or even each individual tumor, will be completely feasible in the near future. The genomic, transcriptomic, epigenomic, and even proteomic and metabolomic changes in various cancers can be viewed as heterogeneous molecular phenotypes of the cancer cells. Many of these changes might be by-products resulting from genome instability or transcriptional and metabolic dysregulation and thus reflect the state of the underlying molecular and metabolic networks. Among these changes, some are necessary for the cancer cells to overcome multiple checkpoints and surveillance mechanisms and expand through clonal selection and expansion, and therefore ultimately enable invasive growth.[10]

There are 2 key points regarding cancer diagnosis and prognosis that should be addressed in the near future: (1) How to sift through numerous multilayer changes within the molecular network of cancer and find critical steps driving the cancer development and metastasis; and (2) How to find essential controllers, better interpret the hallmarks of cancer, and design successful treatment strategies. Toward these goals, both experimental and computational approaches should be investigated to annotate the multilayer cancer network. By taking advantage of the ever-increasing rate and reduced cost of accumulating data, more efforts should be made to achieve the ultimate goal of personal cancer genomics and individualized cancer treatment.

Recent research based on TCGA projects, such as the Pan Cancer Project,[114] have started to integrate cancer types and offer a comprehensive set of cancer systems biology data and new tools for cancer genomics and bioinformatics analysis. In addition to clinical classification of different tumors, this will help to repurpose targeted therapies for cancers under the direction of their molecular pathologies.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Funding

This work was supported by grants from the China Natural National Science Foundation (Grant #91329302, 31210103916, and 91019019), Chinese Ministry of Science and Technology (Grant #2011CB504206), Chinese Academy of Sciences (Grant #KSCX2-EW-R-02 and KSCX2-EW-J-15), and stem cell leading project XDA01010303 to J.D.J.H and by the China Postdoctoral Science Foundation (Grant #2014M551466) to S.J.H.

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA: Cancer J Clin 2013; 63:11-30; PMID: 23335087
2. Edwards BK, Noone A-M, Mariotto AB, Simard EP, Boscoe FP, Henley SJ, Jemal A, Cho H, Anderson RN, Kohler BA, et al. Annual Report to the Nation on the status of cancer, 1975–2010, featuring prevalence of comorbidity and impact on survival among persons with lung, colorectal, breast, or prostate cancer. Cancer 2014; 120:1290-314; PMID:24343171; http://dx.doi.org/10.1002/cncr.28509
3. Smith BD, Smith GL, Hurria A, Hortobagyi GN, Buchholz TA. Future of cancer incidence in the United States: burdens upon an aging, changing nation. J Clin Oncol: Off J Am Soc Clin Oncol 2009; 27:2758-65; PMID:19403886; http://dx.doi.org/10.1200/JCO.2008.20.8983
4. Duque M, Modlin IM, Gupta A, Saif MW. Biomarkers in neuroendocrine tumors. JOP: J Pancreas 2013; 14:372-6; PMID:23846930
5. Chibon F. Cancer gene expression signatures – the rise and fall? Eur J Cancer 2013; 49:2000-9; PMID: 23498875; http://dx.doi.org/10.1016/j.ejca.2013.02.021
6. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, Masson G, Jakobsdottir M, Thorlacius S, Helgason A, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genetic 2007; 39:865-9; PMID:17529974; http://dx.doi.org/10.1038/ng2064
7. Berdasco M, Esteller M. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. Dev Cell 2010; 19:698-711; PMID:21074720; http://dx.doi.org/10.1016/j.devcel.2010.10.005
8. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999; 286:531-7; PMID:10521349; http://dx.doi.org/10.1126/science.286.5439.531
9. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002; 415:530-6; PMID:11823860; http://dx.doi.org/10.1038/415530a
10. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011; 144:646-74; PMID: 21376230; http://dx.doi.org/10.1016/j.cell.2011.02.013
11. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Edgar R. NCBI GEO: mining tens of millions of expression profiles—database and tools update. Nucleic Acids Res 2007; 35:D760-D5; PMID: 17099226; http://dx.doi.org/10.1093/nar/gkl887
12. Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, Holloway E, Kapushesky M, Kemmeren P, Lara GG. ArrayExpress—a public repository for microarray gene expression data at the EBI. Nucleic Acids Res 2003; 31:68-71; PMID: 12519949; http://dx.doi.org/10.1093/nar/gkg091
13. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 2013; 6:pl1; PMID:23550210; http://dx.doi.org/10.1126/scisignal.2004088
14. Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, Menzies A, Teague JW, Futreal PA, Stratton MR. The Catalogue of Somatic Mutations in Cancer (COSMIC). Curr Protoc Hum Genet / Editorial Board, Jonathan L Haines ; et al 2008; Chapter 10:Unit 10 1; PMID:18428421
15. Shepherd R, Forbes SA, Beare D, Bamford S, Cole CG, Ward S, Bindal N, Gunasekaran P, Jia M, Kok CY, et al. Data mining using the Catalogue of Somatic Mutations in Cancer BioMart. Database: J Biol Dat Curation 2011; 2011:bar018; PMID: 21609966; http://dx.doi.org/10.1093/database/bar018
16. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia (New York, NY) 2004; 6:1; PMID:15068665; http://dx.doi.org/10.1016/S1476-5586(04)80047-2
17. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, et al. The landscape of somatic copy-number alteration across human cancers. Nature 2010; 463:899-905; PMID:20164920; http://dx.doi.org/10.1038/nature08822
18. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol 2011; 12:R41; PMID:21527027; http://dx.doi.org/10.1186/gb-2011-12-4-r41
19. Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Tamborero D, Schroeder MP, Jene-Sanz A, Santos A, Lopez-Bigas N. IntOGen-mutations identifies cancer drivers across tumor types. Nat Meth 2013; 10:1081-2; PMID:24037244; http://dx.doi.org/10.1038/nmeth.2642
20. Go H, Jeon YK, Park HJ, Sung SW, Seo JW, Chung DH. High MET gene copy number leads to shorter survival in patients with non-small cell lung cancer. J Thoracic Oncol 2010; 5:305-13; PMID:20107422; http://dx.doi.org/10.1097/JTO.0b013e3181ce3d1d
21. Zhao H, Ljungberg B, Grankvist K, Rasmuson T, Tibshirani R, Brooks JD. Gene expression profiling predicts survival in conventional renal cell carcinoma. PLoS Med 2006; 3:e13; PMID:16318415; http://dx.doi.org/10.1371/journal.pmed.0030013
22. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, et al. A gene-expression signature as a predictor of survival in breast cancer. New Engl J Med 2002; 347:1999-2009; PMID:12490681; http://dx.doi.org/10.1056/NEJMoa021967
23. Di Leva G, Croce CM. miRNA profiling of cancer. Curr Opin Genet Dev 2013; 23:3-11; PMID: 23465882; http://dx.doi.org/10.1016/j.gde.2013.01.004
24. Schoof CR, Botelho EL, Izzotti A, Vasques Ldos R. MicroRNAs in cancer treatment and prognosis. Am J Cancer Res 2012; 2:414-33; PMID:22860232
25. Wulfkuhle JD, Liotta LA, Petricoin EF. Proteomic applications for the early detection of cancer. Nat Rev Cancer 2003; 3:267-75; PMID:12671665; http://dx.doi.org/10.1038/nrc1043
26. Hanash SM, Pitteri SJ, Faca VM. Mining the plasma proteome for cancer biomarkers. Nature 2008; 452:571-9; PMID:18385731; http://dx.doi.org/10.1038/nature06916
27. Marks LS, Fradet Y, Lim Deras I, Blase A, Mathis J, Aubin SM, Cancio AT, Desaulniers M, Ellis WJ, Rittenhouse H. PCA3 molecular urine assay for prostate cancer in men undergoing repeat biopsy. Urology 2007; 69:532-5; PMID:17382159; http://dx.doi.org/10.1016/j.urology.2006.12.014
28. Liu XH, Liu ZL, Sun M, Liu J, Wang ZX, De W. The long non-coding RNA HOTAIR indicates a poor prognosis and promotes metastasis in non-small cell lung cancer. BMC Cancer 2013; 13:464; PMID: 24103700; http://dx.doi.org/10.1186/1471-2407-13-464
29. Geng YJ, Xie SL, Li Q, Ma J, Wang GY. Large intervening non-coding RNA HOTAIR is associated with hepatocellular carcinoma progression. J Int Med Res 2011; 39:2119-28; PMID:22289527; http://dx.doi.org/10.1177/147323001103900608
30. Zhou Y, Zhang X, Klibanski A. MEG3 noncoding RNA: a tumor suppressor. J Mol Endocrinol 2012; 48:R45-53; PMID:22393162; http://dx.doi.org/10.1530/JME-12-0008
31. Wang P, Ren Z, Sun P. Overexpression of the long non-coding RNA MEG3 impairs in vitro glioma cell proliferation. J Cell Biochem 2012; 113:1868-74; PMID:22234798; http://dx.doi.org/10.1002/jcb.24055
32. Heyn H, Esteller M. DNA methylation profiling in the clinic: applications and challenges. Nat Rev Genet 2012; 13:679-92; PMID:22945394; http://dx.doi.org/10.1038/nrg3270
33. Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, Whitaker JW, Tian S, Hawkins RD, Leung D. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell 2013; 153:1134-48; PMID:23664764; http://dx.doi.org/10.1016/j.cell.2013.04.022
34. Stirzaker C, Taberlay PC, Statham AL, Clark SJ. Mining cancer methylomes: prospects and challenges.

Trend Genet 2014; 30:75-84; PMID:24368016; http://dx.doi.org/10.1016/j.tig.2013.11.004

35. Fass L. Imaging and cancer: a review. Mol Oncol 2008; 2:115-52; PMID:19383333; http://dx.doi.org/10.1016/j.molonc.2008.04.001

36. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging 2013; 26:1045-57; PMID:23884657; http://dx.doi.org/10.1007/s10278-013-9622-7

37. Team TNLSTR. Reduced lung-cancer mortality with low-dose computed tomographic screening. New Engl J Med 2011; 365:395-409; PMID:21714641; http://dx.doi.org/10.1056/NEJMoa1102873

38. Kramer BS, Gohagan J, Prorok PC, Smart C. A National Cancer Institute sponsored screening trial for prostatic, lung, colorectal, and ovarian cancers. Cancer 1993; 71:589-93; PMID:8420681; http://dx.doi.org/10.1002/cncr.2820710215

39. Melo FDSE, Vermeulen L, Fessler E, Medema JP. Cancer heterogeneity—a multifaceted view. EMBO Rep 2013; 14:686-95; PMID:23846313; http://dx.doi.org/10.1038/embor.2013.92

40. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell 2010; 17:98-110; PMID:20129251; http://dx.doi.org/10.1016/j.ccr.2009.12.020

41. Erichsen HC, Chanock SJ. SNPs in cancer research and treatment. Brit J Cancer 2004; 90:747-51; PMID:14970847; http://dx.doi.org/10.1038/sj.bjc.6601574

42. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Research 2014, Vol. 42 (Database issue): D1001-D1006

43. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81:559-75; PMID:17701901; http://dx.doi.org/10.1086/519795

44. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet 2007; 39:906-13; PMID:17572673; http://dx.doi.org/10.1038/ng2088

45. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, et al. An integrated genomic analysis of human glioblastoma multiforme. Science 2008; 321:1807-12; PMID:18772396; http://dx.doi.org/10.1126/science.1164382

46. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, Wang D, Masys DR, Roden DM, Crawford DC. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics 2010; 26:1205-10; PMID:20335276; http://dx.doi.org/10.1093/bioinformatics/bAQ126

47. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005; 310:644-8; PMID:16254181; http://dx.doi.org/10.1126/science.1117679

48. MacDonald JW, Ghosh D. COPA–cancer outlier profile analysis. Bioinformatics 2006; 22:2950-1; PMID:16895932; http://dx.doi.org/10.1093/bioinformatics/btl433

49. Tibshirani R, Hastie T. Outlier sums for differential gene expression analysis. Biostatistics 2007; 8:2-8; PMID:16702229; http://dx.doi.org/10.1093/biostatistics/kxl005

50. Wu B. Cancer outlier differential gene expression detection. Biostatistics 2007; 8:566-75; PMID:17021278; http://dx.doi.org/10.1093/biostatistics/kxl029

51. Lian H. MOST: detecting cancer differential gene expression. Biostatistics 2008; 9:411-8; PMID:18048648; http://dx.doi.org/10.1093/biostatistics/kxm042

52. Nguyen-Dumont T, Jordheim LP, Michelon J, Forey N, McKay-Chopin S, Kathleen Cuningham Foundation Consortium for Research into Familial Aspects of Breast C, Sinilnikova O, Le Calvez-Kelm F, Southey MC, Tavtigian SV, et al. Detecting differential allelic expression using high-resolution melting curve analysis: application to the breast cancer susceptibility gene CHEK2. BMC Med Genomics 2011; 4:39; PMID:21569354; http://dx.doi.org/10.1186/1755-8794-4-39

53. Chen J, Zhang D, Zhang W, Tang Y, Yan W, Guo L, Shen B. Clear cell renal cell carcinoma associated microRNA expression signatures identified by an integrated bioinformatics analysis. J Translational Med 2013; 11:169; PMID:23841900; http://dx.doi.org/10.1186/1479-5876-11-169

54. Storey J, Tibshirani R. SAM Thresholding and False Discovery Rates for Detecting Differential Gene Expression in DNA Microarrays. In: Parmigiani G, Garrett E, Irizarry R, Zeger S, eds. The Analysis of Gene Expression Data: Springer New York, 2003:272-90.

55. Smyth GK. Limma: Linear Models for Microarray Data. In: Gentleman R, Carey V, Huber W, Irizarry R, Dudoit S, eds. Bioinformatics and Computational Biology Solutions Using R and Bioconductor: Springer New York, 2005:397-420.

56. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010; 26:139-40; PMID:19910308; http://dx.doi.org/10.1093/bioinformatics/btp616

57. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 2010; 11: R106; PMID:20979621; http://dx.doi.org/10.1186/gb-2010-11-10-r106

58. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol 2013; 31:46-53; PMID:23222703; http://dx.doi.org/10.1038/nbt.2450

59. Li Z, Zhang W, Wu M, Zhu S, Gao C, Sun L, Zhang R, Qiao N, Xue H, Hu Y, et al. Gene expression-based classification and regulatory networks of pediatric acute lymphoblastic leukemia. Blood 2009; 114:4486-93; PMID:19755675; http://dx.doi.org/10.1182/blood-2009-04-218123

60. Jones DR. Meta-analysis: weighing the evidence. Stat Med 1995; 14:137-49; PMID:7754262; http://dx.doi.org/10.1002/sim.4780140206

61. Nordmann AJ, Kasenda B, Briel M. Meta-analyses: what they can and cannot do. Swiss Med Wkly 2012; 142:w13518; PMID:22407741

62. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 2003; 4:249-64; PMID:12925520; http://dx.doi.org/10.1093/biostatistics/4.2.249

63. Brennecke P, Anders S, Kim JK, Kolodziejczyk AA, Zhang X, Proserpio V, Baying B, Benes V, Teichmann SA, Marioni JC, et al. Accounting for technical noise in single-cell RNA-seq experiments. Nat Methods 2013; 10:1093-5; PMID:24056876; http://dx.doi.org/10.1038/nmeth.2645

64. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007; 8:118-27; PMID:16632515; http://dx.doi.org/10.1093/biostatistics/kxj037

65. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K, Irizarry RA. Tackling the widespread and critical impact of batch effects in high-throughput data. Nature Reviews Genetics 2010; 11:733-9; PMID:20838408; http://dx.doi.org/10.1038/nrg2825

66. Teschendorff AE, Zhuang J, Widschwendter M. Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. Bioinformatics 2011; 27:1496-505; PMID:21471010; http://dx.doi.org/10.1093/bioinformatics/btr171

67. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998; 95:14863-8; PMID:9843981; http://dx.doi.org/10.1073/pnas.95.25.14863

68. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, et al. Molecular portraits of human breast tumours. Nature 2000; 406:747-52; PMID:10963602; http://dx.doi.org/10.1038/35021093

69. Qin LX. An integrative analysis of microRNA and mRNA expression–a case study. Cancer Inform 2008; 6:369-79; PMID:19259417

70. Lee H, Kong SW, Park PJ. Integrative analysis reveals the direct and indirect interactions between DNA copy number aberrations and gene expression changes. Bioinformatics 2008; 24:889-96; PMID:18263644; http://dx.doi.org/10.1093/bioinformatics/btn034

71. Shen RL, Olshen AB, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. Bioinformatics 2009; 25:2906-12; PMID:19759197; http://dx.doi.org/10.1093/bioinformatics/btp543

72. Mo Q, Wang S, Seshan VE, Olshen AB, Schultz N, Sander C, Powers RS, Ladanyi M, Shen R. Pattern discovery and cancer gene identification in integrated cancer genomic data. Proc Natl Acad Sci 2013; 110:4245-50; PMID:23431203; http://dx.doi.org/10.1073/pnas.1208949110

73. Yuan Y, Savage RS, Markowetz F. Patient-specific data fusion defines prognostic cancer subtypes. PLoS Comput Biol 2011; 7:e1002227; PMID:22028636; http://dx.doi.org/10.1371/journal.pcbi.1002227

74. Kirk P, Griffin JE, Savage RS, Ghahramani Z, Wild DL. Bayesian correlated clustering to integrate multiple data sets. Bioinformatics 2012; 28:3290-7; PMID:23047558; http://dx.doi.org/10.1093/bioinformatics/bts595

75. Lock EF, Hoadley KA, Marron JS, Nobel AB. Joint and individual variation explained (jive) for integrated analysis of multiple data types. Ann App Stat 2013; 7:523-42; PMID:23745156; http://dx.doi.org/10.1214/12-AOAS597

76. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, Haibe-Kains B, Goldenberg A. Similarity network fusion for aggregating data types on a genomic scale. Nat Methods 2014; 11:333-7; PMID:24464287; http://dx.doi.org/10.1038/nmeth.2810

77. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012; 486:346-52; PMID:22522925

78. Network CGAR. Integrated genomic characterization of endometrial carcinoma. Nature 2013; 497:67-73; PMID:23636398; http://dx.doi.org/10.1038/nature12113

79. Zhang W, Liu Y, Sun N, Wang D, Boyd-Kirkup J, Dou X, Han JD. Integrating genomic, epigenomic, and transcriptomic features reveals modular signatures underlying poor prognosis in ovarian cancer. Cell Rep 2013; 4:542-53; PMID:23933257; http://dx.doi.org/10.1016/j.celrep.2013.07.010
80. Cancer Genome Atlas Research N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 2008; 455:1061-8; PMID:18772890; http://dx.doi.org/10.1038/nature07385
81. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet 2004; 5:101-13; PMID:14735121; http://dx.doi.org/10.1038/nrg1272
82. Han JD. Understanding biological functions through molecular networks. Cell Res 2008; 18:224-37; PMID:18227860; http://dx.doi.org/10.1038/cr.2008.16
83. Saijo N. Critical comments for roles of biomarkers in the diagnosis and treatment of cancer. Cancer Treat Rev 2012; 38:63-7; PMID:21652149; http://dx.doi.org/10.1016/j.ctrv.2011.02.004
84. Itadani H, Mizuarai S, Kotani H. Can systems biology understand pathway activation? Gene expression signatures as surrogate markers for understanding the complexity of pathway activation. Curr Genomics 2008; 9:349-60; PMID:19517027; http://dx.doi.org/10.2174/138920208785133235
85. Papin JA, Hunter T, Palsson BO, Subramaniam S. Reconstruction of cellular signalling networks and analysis of their properties. Nat Rev Mol Cell Biol 2005; 6:99-111; PMID:15654321; http://dx.doi.org/10.1038/nrm1570
86. Newman ME, Girvan M. Finding and evaluating community structure in networks. Physical Rev E Stat Nonlin Soft Mat Phys 2004; 69:026113; PMID:14995526; http://dx.doi.org/10.1103/PhysRevE.69.026113
87. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature 1998; 393:440-2; PMID:9623998; http://dx.doi.org/10.1038/30918
88. Cui Q, Ma Y, Jaramillo M, Bari H, Awan A, Yang S, Zhang S, Liu L, Lu M, O'Connor-McCourt M, et al. A map of human cancer signaling. Mol Syst Biol 2007; 3:152; PMID:18091723; http://dx.doi.org/10.1038/msb4100200
89. Segre D, Deluna A, Church GM, Kishony R. Modular epistasis in yeast metabolism. Nat Genet 2005; 37:77-83; PMID:15592468
90. Xue H, Xian B, Dong D, Xia K, Zhu S, Zhang Z, Hou L, Zhang Q, Zhang Y, Han JD. A modular network model of aging. Mol Syst Biol 2007; 3:147; PMID:18059442; http://dx.doi.org/10.1038/msb4100189
91. Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. Mol Syst Biol 2007; 3:140; PMID:17940530; http://dx.doi.org/10.1038/msb4100180
92. Kind T, Liu KH, Lee do Y, DeFelice B, Meissen JK, Fiehn O. LipidBlast in silico tandem mass spectrometry database for lipid identification. Nat Methods 2013; 10:755-8; PMID:23817071; http://dx.doi.org/10.1038/nmeth.2551
93. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering of regulatory networks in human B cells. Nat Genet 2005; 37:382-90; PMID:15778709; http://dx.doi.org/10.1038/ng1532
94. Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao X, Snyder EY, Sulman EP, Anne SL, Doetsch F, Colman H, et al. The transcriptional network for mesenchymal transformation of brain tumours. Nature 2009; 463:318-25; PMID:20032975; http://dx.doi.org/10.1038/nature08712
95. Mani KM, Lefebvre C, Wang K, Lim WK, Basso K, Dalla-Favera R, Califano A. A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. Mol Syst Biol 2008; 4:169; PMID:18277385; http://dx.doi.org/10.1038/msb.2008.2
96. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003; 13:2498-504; PMID:14597658; http://dx.doi.org/10.1101/gr.1239303
97. Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, Pico AR, Bader GD, Ideker T. A travel guide to Cytoscape plugins. Nat Methods 2012; 9:1069-76; PMID:23132118; http://dx.doi.org/10.1038/nmeth.2212
98. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B. Integration of biological networks and gene expression data using Cytoscape. Nat Protoc 2007; 2:2366-82; PMID:17947979; http://dx.doi.org/10.1038/nprot.2007.324
99. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000; 25:25-9; PMID:10802651; http://dx.doi.org/10.1038/75556
100. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 2012; 40:D109-14; PMID:22080510; http://dx.doi.org/10.1093/nar/gkr988
101. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, Ami GOH, Web Presence Working G. AmiGO: online access to ontology and annotation data. Bioinformatics 2009; 25:288-9; PMID:19033274; http://dx.doi.org/10.1093/bioinformatics/btn615
102. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 2005; 21:3448-9; PMID:15972284; http://dx.doi.org/10.1093/bioinformatics/bti551
103. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009; 4:44-57; PMID:19131956; http://dx.doi.org/10.1038/nprot.2008.211
104. Gundem G, Perez-Llamas C, Jene-Sanz A, Kedzierska A, Islam A, Deu-Pons J, Furney SJ, Lopez-Bigas N. IntOGen: integration and data mining of multidimensional oncogenomic data. Nat Methods; 7:92-3; PMID:20111033; http://dx.doi.org/10.1038/nmeth0210-92
105. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005; 102:15545-50; PMID:16199517; http://dx.doi.org/10.1073/pnas.0506580102
106. Kim SY, Volsky DJ. PAGE: parametric analysis of gene set enrichment. BMC Bioinformatics 2005; 6:144; PMID:15941488; http://dx.doi.org/10.1186/1471-2105-6-144
107. Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. J Comput Biol 2000; 7:601-20; PMID:11108481; http://dx.doi.org/10.1089/106652700750050961
108. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP. Causal protein-signaling networks derived from multiparameter single-cell data. Science 2005; 308:523-9; PMID:15845847; http://dx.doi.org/10.1126/science.1105809
109. Needham CJ, Bradford JR, Bulpitt AJ, Westhead DR. Inference in Bayesian networks. Nat Biotechnol 2006; 24:51-3; PMID:16404397; http://dx.doi.org/10.1038/nbt0106-51
110. Verduijn M, Peek N, Rosseel PM, de Jonge E, de Mol BA. Prognostic Bayesian networks I: rationale, learning procedure, and clinical use. J Biomed Infor 2007; 40:609-18; PMID:17704008; http://dx.doi.org/10.1016/j.jbi.2007.07.003
111. Gevaert O, De Smet F, Timmerman D, Moreau Y, De Moor B. Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. Bioinformatics 2006; 22:e184-90; PMID:16873470; http://dx.doi.org/10.1093/bioinformatics/btl230
112. Sesen MB, Nicholson AE, Banares-Alcantara R, Kadir T, Brady M. Bayesian networks for clinical decision support in lung cancer care. PloS One 2013; 8:e82349; PMID:24324773; http://dx.doi.org/10.1371/journal.pone.0082349
113. Chickering DM. A transformational characterization of equivalent bayesian network structures In Proceedings of Eleventh Conference on Uncertainty in Artificial Intelligence, Montreal, QC, p. 87–98. Morgan Kaufmann Publishers.
114. Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM, Network CGAR. The cancer genome atlas pan-cancer analysis project. Nat Genet 2013; 45:1113-20; PMID:24071849; http://dx.doi.org/10.1038/ng.2764