



Mechanism-Centric Approaches for Biomarker Detection and Precision Therapeutics in Cancer

Christina Y. Yu¹ and Antonina Mitrofanova^{1,2*}

¹ Department of Biomedical and Health Informatics, School of Health Professions, Rutgers, The State University of New Jersey, Newark, NJ, United States, ² Rutgers Cancer Institute of New Jersey, Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

OPEN ACCESS

Edited by:

Rosalba Giugno,
University of Verona, Italy

Reviewed by:

Vittorio Fortino,
University of Eastern Finland, Finland
Sailu Yellaboina,
CR Rao Advanced Institute
of Mathematics, Statistics
and Computer Science, India

*Correspondence:

Antonina Mitrofanova
amitrofa@shp.rutgers.edu

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

Received: 30 March 2021

Accepted: 28 June 2021

Published: 02 August 2021

Citation:

Yu CY and Mitrofanova A (2021)
Mechanism-Centric Approaches
for Biomarker Detection and Precision
Therapeutics in Cancer.
Front. Genet. 12:687813.
doi: 10.3389/fgene.2021.687813

Biomarker discovery is at the heart of personalized treatment planning and cancer precision therapeutics, encompassing disease classification and prognosis, prediction of treatment response, and therapeutic targeting. However, many biomarkers represent passenger rather than driver alterations, limiting their utilization as functional units for therapeutic targeting. We suggest that identification of driver biomarkers through mechanism-centric approaches, which take into account upstream and downstream regulatory mechanisms, is fundamental to the discovery of functionally meaningful markers. Here, we examine computational approaches that identify mechanism-centric biomarkers elucidated from gene co-expression networks, regulatory networks (e.g., transcriptional regulation), protein–protein interaction (PPI) networks, and molecular pathways. We discuss their objectives, advantages over gene-centric approaches, and known limitations. Future directions highlight the importance of input and model interpretability, method and data integration, and the role of recently introduced technological advantages, such as single-cell sequencing, which are central for effective biomarker discovery and time-cautious precision therapeutics.

Keywords: biomarkers, treatment response, precision medicine, predictive models, mechanism-centric approaches

INTRODUCTION

In the past two decades, the advancement of high-throughput technologies has led to the discovery of genomic, transcriptomic, and epigenomic modalities involved in cancer initiation, progression, and treatment response. Multiple groups have started to effectively utilize molecular data produced by high-throughput oncology experiments to identify biomarkers of progression and therapeutic response in cancer patients (Sorlie et al., 2001; Zhang et al., 2001; van't Veer et al., 2002; Zhan et al., 2002, 2006; Sotiriou et al., 2003; Ayers et al., 2004; Allen et al., 2006; Jain et al., 2009; Lim et al., 2009; Petty et al., 2009; Zhao et al., 2009; Carro et al., 2010; Lefebvre et al., 2010; Shaughnessy et al., 2011; Bae et al., 2013; Aytes et al., 2014, 2018; Mitrofanova et al., 2015; Robinson et al., 2015; Wang et al., 2016; Giulietti et al., 2017; Heng et al., 2017; Hoadley et al., 2018; Abida et al., 2019; Epsi et al., 2019; Arriaga et al., 2020; Panja et al., 2020; Rahem et al., 2020). Yet, our understanding of the mechanisms involving these modalities, their upstream regulation, and effective therapeutic targeting remains incomplete.

A biomarker is an objective measure (e.g., classically a genomic/transcriptomic/epigenomic alteration, gene, protein, metabolite, or their groups), typically used to predict the incidence of disease, its progression, or treatment outcome (Strimbu and Tavel, 2010; McDermott et al., 2013). In the context of oncology, biomarkers are classically used for cancer risk assessment and screening, tumor staging, disease recurrence, selection of initial therapy, alternative therapy choices, and monitoring for therapeutic toxicities (Ludwig and Weinstein, 2005). While employed in clinical use, the existing biomarkers are still sparse and suffer from issues of reproducibility and heterogeneity, alongside a lack of understanding of their underlying regulatory mechanisms (Ludwig and Weinstein, 2005; Boutros, 2015).

One of the reasons for such a knowledge gap is the fact that the majority of biomarkers are identified from *gene-centric* approaches (we will refer to gene/protein/metabolite etc.,-centric approaches as gene-centric approaches for simplicity), where either a specific gene is investigated (based on previous biological assumptions) or a gene(s) is selected based on differential behavior without connection to the upstream and downstream molecular mechanisms. Gene-centric findings are often limited in mechanistic interpretability and connectivity to other molecular processes, positioning such biomarkers as passengers, rather than drivers, of the biological process and thus are often dataset specific (Michiels et al., 2005; Chng et al., 2016).

In classical gene-centric approaches, genes (without their connections to one another or underlying mechanisms) are utilized as inputs into white- and black-box statistical and machine learning models, which have been successfully applied to identify gene-centric markers in breast cancer (van't Veer et al., 2002; Wang et al., 2005; Zhang et al., 2013), lung cancer (Beer et al., 2002), multiple myeloma (Shaughnessy et al., 2007; Kuiper et al., 2012), colon cancer (Zhang et al., 2001; Yan et al., 2012), and prostate cancer (Garzotto et al., 2005; Erho et al., 2013), among many others. It is important to note that in white-box models (e.g., linear regression and decision trees) the relationship between input variables (i.e., genes) and output variables (i.e., disease outcomes) is understandable/explainable as they often identify linear or monotonic relationships (Zhang et al., 2001; Garzotto et al., 2005; Rosenfeld et al., 2008; Huo et al., 2017; Panja et al., 2018). On the other hand, black-box models (e.g., neural networks, gradient boosting, or ensemble models such as random forest) are able to capture non-linear/non-monotonic relationships, yet often suffer from model interpretability and subsequent limited clinical adoption (Wang et al., 2009; Ayer et al., 2010; Zhang et al., 2013). Even though both white- and black-box learning are excellent tools for predictive modeling, they mostly capture associative relationships when applied as gene-centric approaches and often miss the complexity of mechanisms inherent in biological systems, especially in the context of cancer.

Several groups have addressed this problem by developing biomarker discovery methods based on *mechanism-centric* approaches, which are not focused on single genes and take into account complex mechanisms implicated in cancer initiation, progression, and treatment response. In this review, we will

discuss the mechanism-centric approaches based on construction and mining of co-expression networks (Freeman, 1977; Zhang and Horvath, 2005; Zhang and Huang, 2014; Han et al., 2016), regulatory networks (Basso et al., 2005; Lefebvre et al., 2010; Alvarez et al., 2016; Dhingra et al., 2017), protein-protein interaction (PPI) networks (Chuang et al., 2007), and molecular pathways (Epsi et al., 2019; Rahem et al., 2020; **Figure 1**). Through an in-depth understanding of upstream and downstream molecular mechanisms, such techniques open a door for the discovery of functionally interpretable molecular drivers (rather than passengers) and potential targets for precision therapeutics.

MECHANISM-CENTRIC COMPUTATIONAL APPROACHES FOR BIOMARKER DISCOVERY

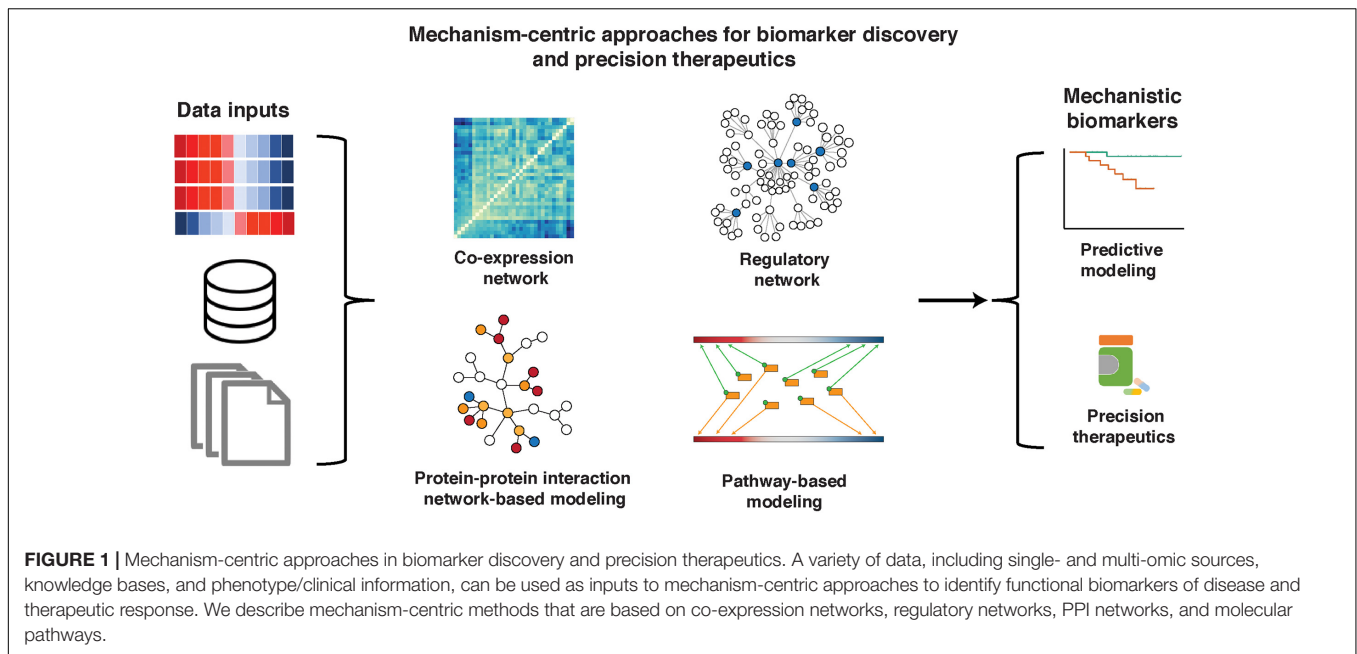
Gene Co-expression Network Analysis

Gene co-expression networks define groups of genes that show similar/related expression patterns across an entire dataset. Highly associated genes are clustered together into modules, with the underlying rationale that co-expressed genes are likely to be co-regulated. We depict two methods, weighted gene co-expression network analysis (WGCNA) (Langfelder and Horvath, 2008) and local maximal Quasi-Clique Merger (ImQCM) (Zhang and Huang, 2014), for network construction and module detection. Identified modules are defined as tightly connected groups of genes (potentially protein/gene complexes), which are then associated with clinical features to determine functionally relevant molecular structures. We also describe methods to mine such co-expression networks that include condition-specific network mining (Han et al., 2016), eigengene association (Alter et al., 2000; Zhang and Horvath, 2005), and network connectivity/hub analysis (Freeman, 1977).

Network Construction: WGCNA and ImQCM

In general, co-expression network construction is based on a similarity matrix that describes the measure of association between a gene to all other genes (the simplest of similarity measures being correlation) (**Figure 2A**). An undirected network is constructed from the similarity matrix and is comprised of nodes denoting genes and edges denoting the associations (e.g., correlation) between genes.

One of the most well-known methods for gene co-expression network reconstruction is WGCNA, which was one of the earliest methods that proposed using weighted networks (**Figure 2B**; Zhang and Horvath, 2005). The advantage of weighted, compared to unweighted, network construction is the ability to assign meaningful weights to relationships/edges, which eliminates a need for threshold assignment and prevents information loss. WGCNA calculates correlation between pairs of genes and transforms the correlation measure into a topological overlap measure in order to minimize effects of noise and spurious associations. The resulting matrix is subjected to hierarchical clustering to determine groups of co-expressed genes, also



referred to as gene modules. An R package for WGCNA is freely available (Langfelder and Horvath, 2008).

Because WGCNA module identification is based on hierarchical clustering, genes cannot be assigned to multiple modules, exposing WGCNA's limitation since many genes participate in multiple biological processes and often perform multiple functions. An alternative weighted co-expression method which allows genes to have multiple co-memberships in different modules is lmQCM (Figure 2C; Zhang and Huang, 2014). The lmQCM algorithm identifies densely connected subnetworks (i.e., quasi-cliques) using a greedy search algorithm which allows module overlaps (Ou and Zhang, 2007). In addition to allowing genes to be assigned to multiple modules, lmQCM can also identify smaller modules, which can highlight more specific and interpretable biological connections as compared to much larger modules of WGCNA that frequently contain over a thousand genes (Zhang and Huang, 2014; Yu et al., 2019). This algorithm is freely available as an R package¹ and a web-tool (Huang et al., 2021).

Network Mining: Centered Concordance Index, Eigengenes, and Hubs

Co-expression networks can be mined to determine the functional significance of their modules or identify functionally relevant genes. Here, we discuss two techniques for module mining [Centered Concordance Index (CCI) (Han et al., 2016) and eigengenes (Alter et al., 2000; Horvath and Dong, 2008)] and two techniques to identify hub genes [intramodular connectivity (Zhang and Horvath, 2005) and betweenness centrality (Freeman, 1977)].

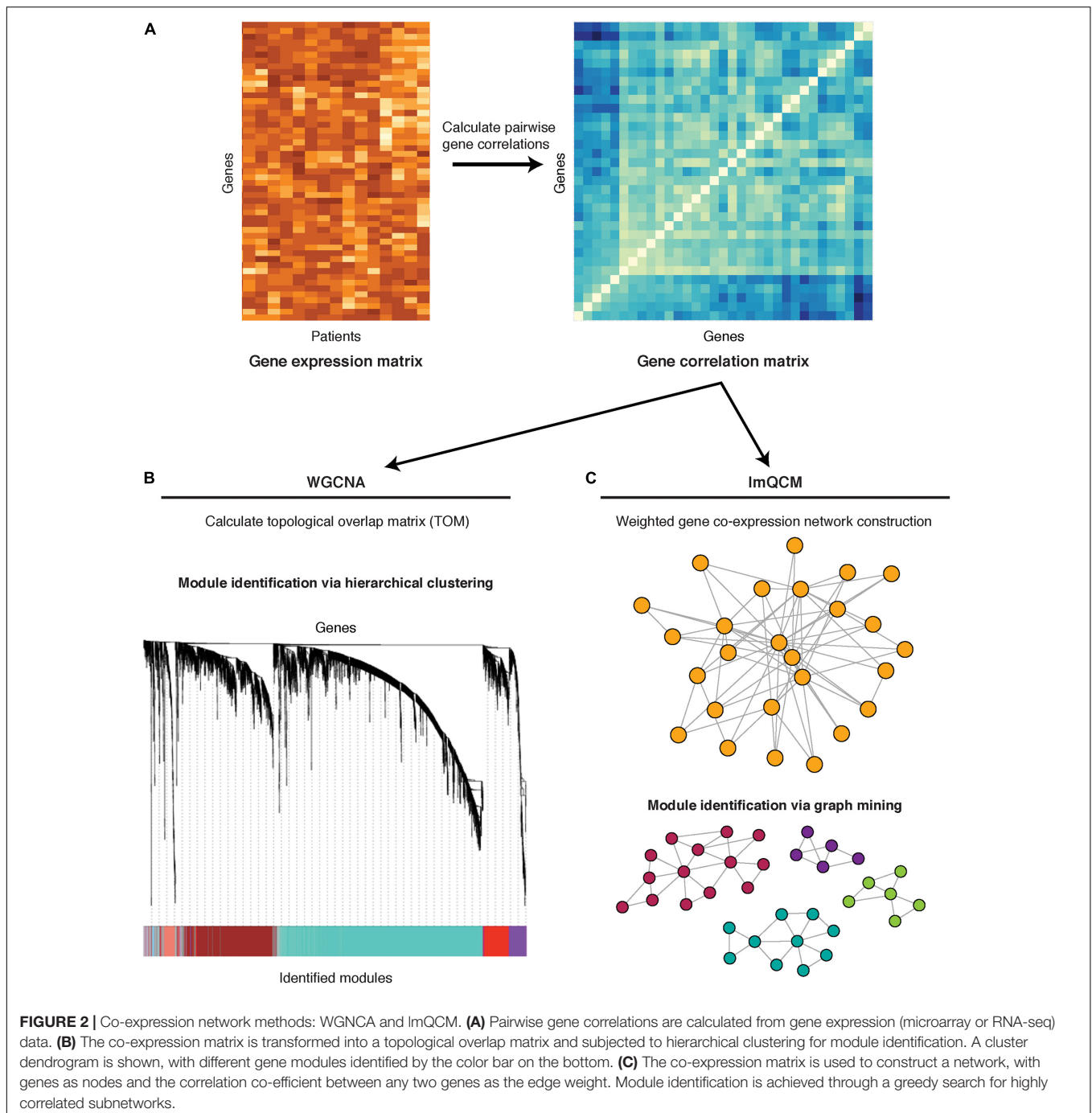
Centered Concordance Index has been developed to identify modules specific to each condition/phenotype. In particular,

the CCI evaluates the concordance of gene expression profiles within a module based on singular value decomposition and is used to identify modules that are highly co-expressed in one condition over another (Han et al., 2016). Han et al. (2016) and Yu et al. (2019), respectively, identified several gene modules specific to lung adenocarcinoma and multiple myeloma precursors compared to non-cancer controls. The CCI is useful in identifying modules specific to phenotype conditions but has yet to be used to associate modules with continuous outcomes.

The eigengene approach transforms modules into weighted vectors, which mathematically correspond to their contribution to the first principal component in principal component analysis (Alter et al., 2000; Horvath and Dong, 2008). Eigengenes are then able to be associated with clinical features (including continuous outcomes) using correlation/association measures. For instance, Liu et al. (2015a) used the eigengene approach to identify two modules significantly associated with poor outcome in ER + breast cancer patients treated with tamoxifen. Liu et al. (2015b) and Zhang J. et al. (2020) associated module eigengenes derived from breast cancer patient data with clinical features such as survival status, tumor metastasis, and chemotherapy response. Han et al. (2019) identified module eigengenes strongly associated with patient survival in neuroblastoma.

The translational applicability of modules can be hampered by their relatively large size and might benefit from identification of hub genes within modules. Several measures have been developed to identify hubs, including intramodular connectivity and betweenness centrality. In particular, intramodular connectivity for gene i is defined as the sum of edge weights between gene i and the other genes in the module (Zhang and Horvath, 2005). Genes with the highest connectivity are considered hub genes and have been shown to play key roles in maintaining essential cellular functions (Jeong et al., 2001) and significantly associated with patient survival in breast cancer (Liu et al., 2015a;

¹<https://cran.r-project.org/package=lmQCM>



Tang et al., 2018; Jia et al., 2020; Tian et al., 2020; Zhang J. et al., 2020), glioblastoma (Horvath et al., 2006; Yang et al., 2018; Tang et al., 2019), hepatocellular carcinoma (Hu et al., 2020; Song et al., 2020), and pancreatic ductal adenocarcinoma (Giulietti et al., 2016), among others. Some of these findings have been experimentally validated, such as the ASPM hub gene in glioblastoma (Horvath et al., 2006) and FAM171A1, NDFIP1, SKP1, and REEP5 hub genes in breast cancer (Tian et al., 2020).

An alternative measure to identify hub genes is betweenness centrality, which is a network topology metric used to identify

central nodes in a graph based on a shortest paths algorithm (Freeman, 1977). The betweenness centrality of gene i is a measure of the number of shortest paths connecting any two genes which pass through i . Genes with the highest betweenness scores are considered hubs and are believed to play an important role in information transfer within the network. For instance, Wang et al. analyzed modules with the betweenness centrality measure to identify eight hub genes that were significantly associated with overall survival in breast cancer patients (Wang C. C. N. et al., 2019).

Regulatory Network Analysis

In recent years, molecular regulatory networks have received much attention from the scientific community due to their ability to capture complexity of molecular interactions present in cancer context-specific tissues (Butte and Kohane, 2000; Butte et al., 2000; Friedman et al., 2000; Basso et al., 2005; Margolin et al., 2006a,b; Werhli and Husmeier, 2007; Huynh-Thu et al., 2010; Lefebvre et al., 2010; Aytes et al., 2014). Regulatory networks define regulatory relationships between regulators (e.g., transcriptional regulators, splicing regulators, post-translational regulators, etc.), and their potential targets (e.g., genes, proteins, etc.). Such regulatory relationships provide key information about upstream and downstream regulations to infer cellular mechanisms for creating potential causal models of disease and outperform co-expression networks in their interpretability and functionally relevant determinants. Several methods have tackled reconstruction of regulatory networks using mutual information (Butte and Kohane, 2000; Basso et al., 2005; Margolin et al., 2006a), Bayesian networks (Friedman et al., 2000; Werhli and Husmeier, 2007), and regression trees (Huynh-Thu et al., 2010), to name a few. Readers are encouraged to consult the following reviews for a comprehensive overview of the different computational underpinnings employed in regulatory network analysis (Markowitz and Spang, 2007; Karlebach and Shamir, 2008; Hecker et al., 2009; Lee and Tzou, 2009; Emmert-Streib et al., 2014). Here, we focus on transcriptional [Algorithm for the Reconstruction of Gene Regulatory Networks (ARACNe) (Margolin et al., 2006a)] and multi-omic [RegNetDriver (Dhingra et al., 2017)] regulatory networks and their mining [i.e., Master Regulator Inference Algorithm (MARINA) (Lefebvre et al., 2010), Virtual Inference of Protein-activity by Enriched Regulon analysis (VIPER) (Alvarez et al., 2016), etc.] in the context of cancer biomarker studies.

Transcriptional Regulatory Networks

The role of transcriptional regulation has been widely studied in cancer, including discovery of MYC (Gabay et al., 2014), Sox2 (Boumahdi et al., 2014), and the FOXO family (Jiramongkol and Lam, 2020) as important players in cancer initiation and progression. Transcriptional regulatory networks depict interactions between transcription factors (TFs)/co-factors (co-TFs) and their transcriptional targets, allowing the study of differential behavior in transcriptional machinery that govern oncogenic process.

Network construction: ARACNe

One of the most known and widely experimentally validated methods for transcriptional network reconstruction is ARACNe (Margolin et al., 2006a,b). This information-theoretic algorithm utilizes tissue-specific gene expression profiles to estimate pairwise mutual information between expression levels of TFs/co-TFs and expression levels of their potential (activated or repressed) targets. The advantage of using mutual information to measure such relationships lies in its ability to measure not only linear (which would be captured for example by the Pearson correlation) or monotonic (which would be captured for

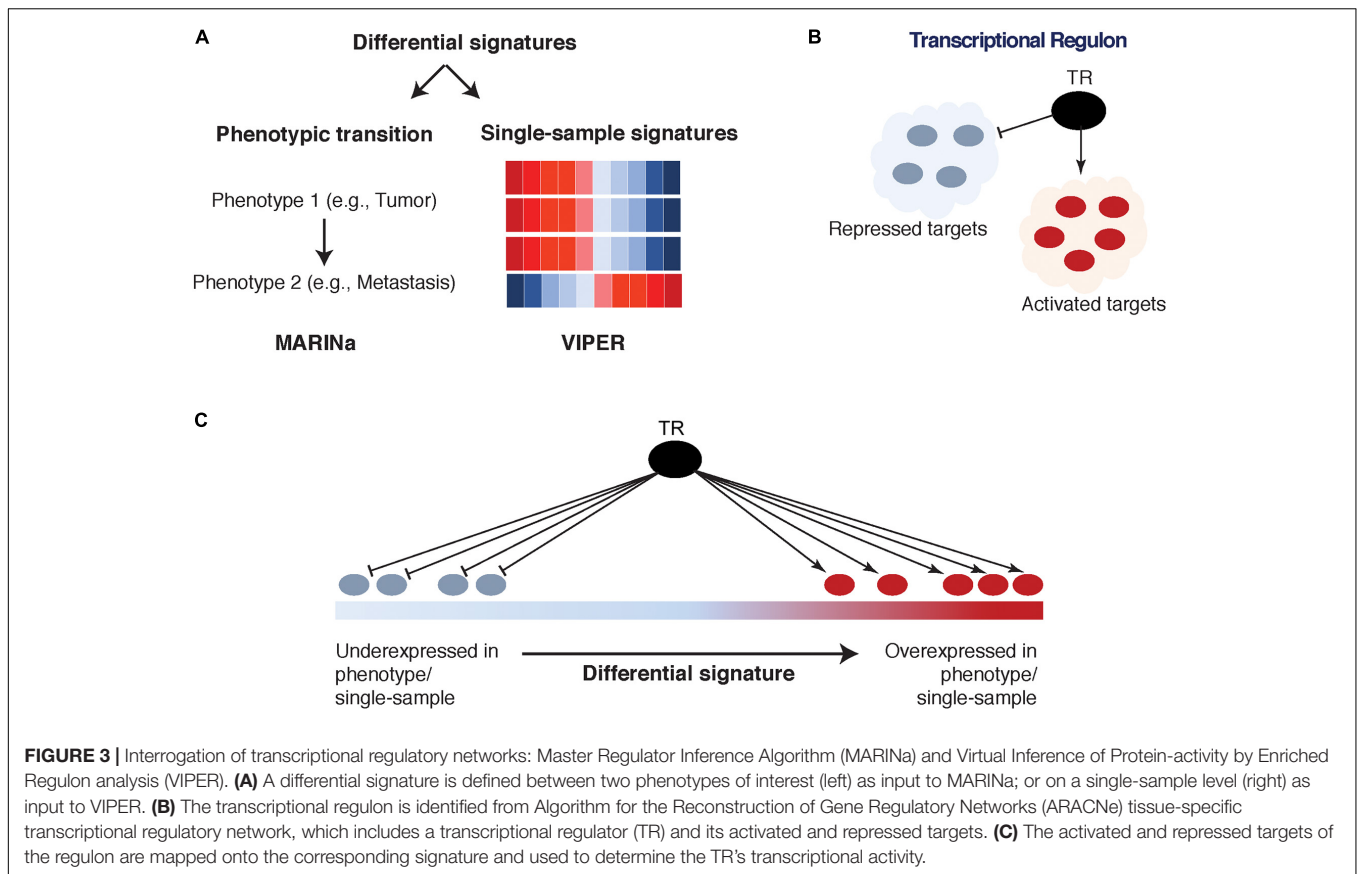
example by Spearman correlation) relationships, but also non-linear associations. Another novelty in transcriptional network reconstruction is introduced by the data processing inequality, which eliminates any “indirect” regulatory relationship through the principle that mutual information on the indirect path cannot exceed mutual information on any part of the direct path. Data processing inequality results in a regulatory network that includes primarily direct TF/co-TF-target interactions. ARACNe has been widely applied to several normal physiological and pathological conditions, including B-cell interactome (Basso et al., 2005), breast cancer (Lim et al., 2009; Remo et al., 2015; Walsh et al., 2017), prostate cancer (Aytes et al., 2014), colorectal cancer (Bae et al., 2013; Cordero et al., 2014; Sanz-Pamplona et al., 2014; Eskandari et al., 2018), glioma (Carro et al., 2010), T-cell acute lymphoblastic leukemia (Palomero et al., 2006), and multiple myeloma (Agnelli et al., 2011), among others. Software for ARACNe is freely available for download.²

Network mining: MARINA and VIPER

The ARACNe network can be effectively interrogated (i.e., mined) using MARINA (Lefebvre et al., 2010) and VIPER (Alvarez et al., 2016), two algorithms that identify TFs/co-TFs as driver biomarkers associated with specific phenotypes (e.g., cancer initiation, cancer progression, metastasis, treatment response, etc.). Specifically, MARINA (Lim et al., 2009; Lefebvre et al., 2010) requires a differentially expressed signature, defined as a ranked list of genes between any two phenotypes of interest. Then, the activated and repressed targets for each TF/co-TF (as inferred by ARACNe) are assessed for their enrichment in the over- and under-expressed parts of this signature (Lefebvre et al., 2010; **Figure 3**). Such enrichment is referred to as TF/co-TF transcriptional activity, and if it is statistically significant, the TF/co-TF is referred to as a Master Regulator (MR). As a result of this analysis, a TF/co-TF is considered an “activated” MR if its activated targets are significantly enriched in the over-expressed part of the signature and/or its repressed targets are significantly enriched in the under-expressed part of the signature. Conversely, a “repressed” MR exhibits the opposite behavior. It is important to note that TF/co-TF transcriptional activity is not defined based on the differential expression of TFs/co-TFs themselves but instead on the differential expression of their transcriptional targets. This allows the identification of TFs/co-TFs that are not necessarily differentially expressed but are modified on the post-translational level and would otherwise be missed by traditional association methods.

Master Regulator Inference Algorithm has successfully identified MRs in various cancers, including prostate cancer (Aytes et al., 2014, 2018; Mitrofanova et al., 2015; Talos et al., 2017), breast cancer (Lim et al., 2009; Fletcher et al., 2013; Remo et al., 2015), pancreatic cancer (Sartor et al., 2014), ovarian cancer (Zhang et al., 2015), glioma (Carro et al., 2010; Sonabend et al., 2014), T cell acute lymphoblastic leukemia (Della Gatta et al., 2012), and diffuse large B cell lymphoma (Ying et al., 2013; Bisikirska et al., 2016). These biomarkers also serve as valuable therapeutic targets and their silencing could potentially

²<http://califano.c2b2.columbia.edu/aracne>



have a significant effect on inhibition of malignant phenotype. To this extent, Mitrofanova et al. developed a computational algorithm to predict drug combinations that inhibit activity levels of FOXM1 and CENPF (MRs in malignant prostate cancer) and demonstrated that their therapeutic inhibition significantly improved cancer course (Mitrofanova et al., 2015). MARINA is freely available for download.³

At the same time, VIPER estimates TF/co-TF transcriptional activity on an individual sample-based level, as opposed to a two-phenotype signature-based level required by MARINA (Alvarez et al., 2016; **Figure 3**). In fact, while MARINA requires carefully selected multiple samples of the same phenotype to construct a differential expression signature, VIPER is able to utilize single-sample analysis by scaling the overall patient cohort (to its average expression for each gene). Furthermore, several advantages of VIPER include estimation of TF/co-TF activity through a so-called mode of regulation (taking into account whether targets are activated, repressed, or their direction cannot be determined), inference of regulator-target interaction confidence, and accounting for target overlap between different regulators (Alvarez et al., 2016). VIPER was shown to accurately infer aberrant oncoprotein activity induced by somatic mutations, across multiple cancer types (Alvarez et al., 2016). An R package is freely available.⁴

³<http://califano.c2b2.columbia.edu/marina>

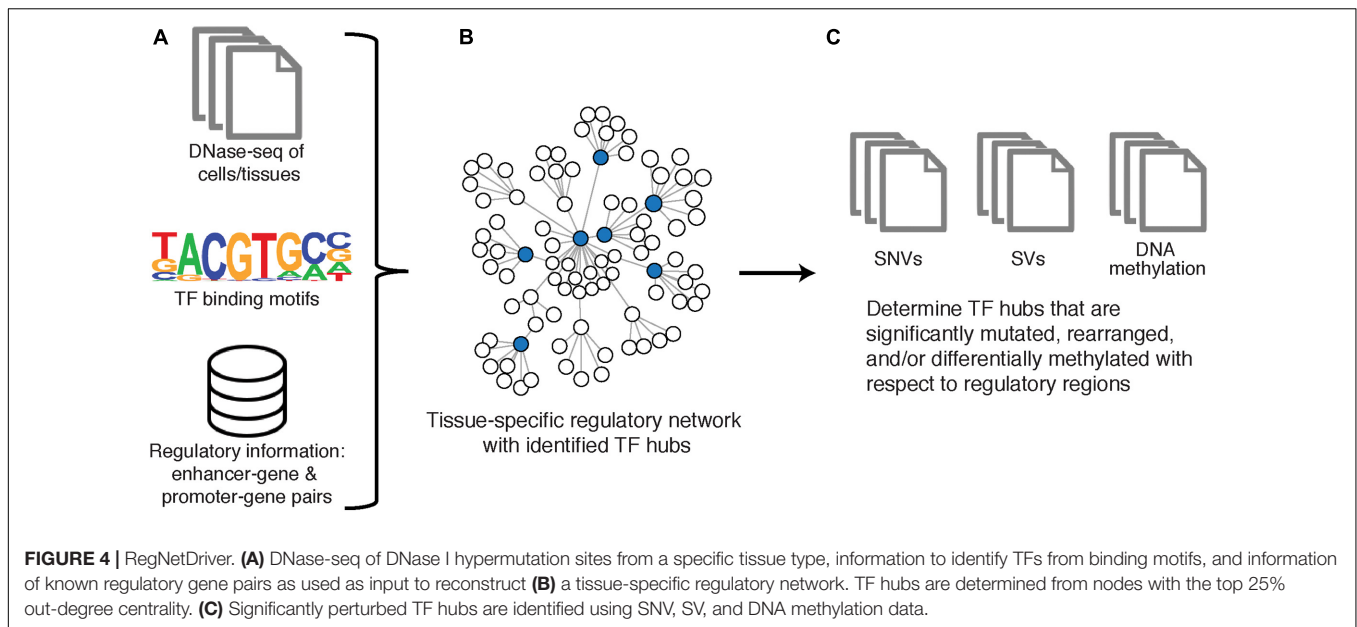
⁴<http://doi.org/10.18129/B9.bioc.viper>

Multi-Omic Regulatory Network

Multi-omic data integration is another avenue to improve interpretability and discovery of functionally relevant biomarkers. Integration of different data modalities can increase the confidence of the overall findings since gene regulation is a complex process affected by multiple factors, such as gene mutations, structural variants, epigenomics, and more.

Network construction: RegNetDriver, step I

RegNetDriver is an algorithm for multi-omic tissue-specific regulatory network construction and analysis (Dhingra et al., 2017; **Figure 4**). The regulatory network reconstructed by RegNetDriver represents a two-layered relationship: (i) connecting TFs to promoter/enhancer regions; and (ii) further connecting promoter/enhancer regions to their corresponding target genes. To reconstruct relationships between TFs and promoters/enhancers of potential targets, Dhingra et al. utilize tissue-specific (i.e., prostate epithelium) DNase I hypersensitive sites to define accessible regulatory DNA regions and integrate this information with promoter/enhancer annotations from ENCODE (Encode Project Consortium, 2012) and GENCODE (Harrow et al., 2012). TFs are then connected to promoters/enhancers based on the enrichment of their binding motifs. Promoters/enhancers are further connected to their target genes through significant correlation of promoter/enhancer region activity signals (estimated using bisulfite sequencing and ChIP-seq data) with target gene



expression profiles (estimated using RNA-seq data). Note that this is a directed two-layered network that estimates relationships between TFs and their transcriptional targets through their corresponding promoter/enhancer associations.

Network mining: RegNetDriver, step II

This network is then utilized to identify TF hubs with genomic and epigenomic alterations that can potentially cause large perturbations in this tissue-specific network. Specifically, TFs are first mined on degree centrality, such that the top 25% of TFs with the greatest number of outgoing edges are defined as hubs. Next, to identify TF hubs significantly affected on genomic and epigenomic levels in prostate cancer, they are evaluated for the presence of prostate-cancer specific genomic alterations (single nucleotide variants and structural variants) and DNA methylation changes in their coding and non-coding regulatory regions. In Dhingra et al., RegNetDriver nominated three TFs as regulatory drivers in prostate cancer, with functional validation conducted on *ERF* (Dhingra et al., 2017). RegNetDriver is freely available for download.⁵

Protein–Protein Interaction Network-Based Analysis

Another important avenue in mechanism-centric biomarker discovery is PPIs. Such interactions elucidate putative protein complexes, which are known to perform critical functions within the cell and include for example the pre-initiation complex for RNA transcription (Greber and Nogales, 2019), the spliceosome for pre-mRNA splicing (Chen et al., 2007), and the ribosome for translation of mRNA to protein (Wilson and Doudna Cate, 2012), among others. Cancer cells in particular have been shown to deregulate protein complexes for their sustained proliferation, survival, and metastasis (Robichaud et al., 2019). In recent years,

numerous public databases have cataloged networks of known and predicted PPIs, such as STRING (Szklarczyk et al., 2019), IntAct (Orchard et al., 2014), CellCircuits (Mak et al., 2007), and PINA (Cowley et al., 2012) [more comprehensive lists are described by Huang et al. (2018) and Miryala et al. (2018)]. Here, we describe the method from Chuang et al. (2007), which effectively combines PPI networks with gene expression data and evaluates these hybrid subnetworks as mechanism-centric biomarkers of breast cancer metastasis (Figure 5).

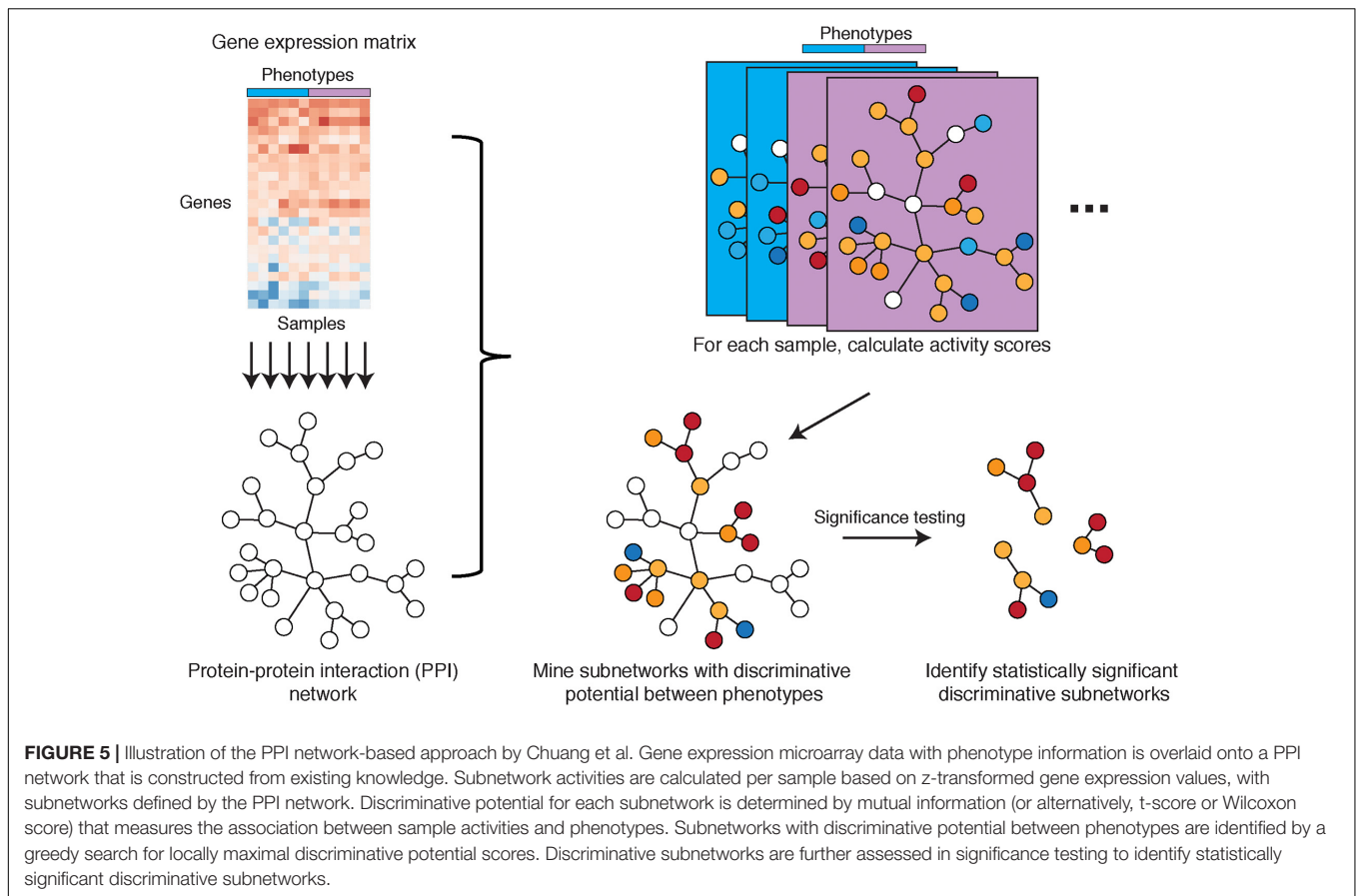
Network Construction: Chuang et al., Step I

Chuang et al. introduce a hybrid approach to combine a PPI network with tissue-specific gene expression profiles across patient samples. The PPI network is comprised of nodes representing proteins and edges representing a characterized PPI, utilizing subnetworks from CellCircuits. Tissue-specific gene expression data are then overlaid onto all PPI subnetworks. For each subnetwork, its activity in each sample/patient is defined as a combination of z-scores for the subnetwork genes. This defines patient-specific vectors of subnetwork activities, which are then mined for phenotype associations.

Network Mining: Chuang et al., Step II

Activities of subnetworks are evaluated for their association with specific phenotypes (e.g., metastatic and non-metastatic), where associations can be calculated by mutual information, t-score, or Wilcoxon score and is referred to as the subnetwork discriminative potential/score. Next, the method selects subnetworks with a locally maximal discriminative score and performs significance testing to ensure subnetworks are non-random and robust. In classification performance on a test cohort, the authors found that the subnetwork markers identified using this PPI network-based approach showed higher AUC in classifying metastatic versus non-metastatic samples compared to single-gene markers, random subnetworks, and gene sets

⁵<https://khouranalab.med.cornell.edu/RegNetDriver.html>



from other annotation databases such as GO and MSigDB. Importantly, the method by Chuang et al. showed better biomarker reproducibility (i.e., higher overlap between markers) between two different breast cancer studies, outperforming gene-centric methods (Chuang et al., 2007).

Pathway-Based Analysis: pathCHEMO and pathER

Recently, pathway-based biomarker algorithms, such as pathCHEMO (Epsi et al., 2019) and pathER (Rahem et al., 2020), have demonstrated that discovery approaches that encompass information from biological pathways significantly outperform gene-centric methods which do not take into account pathway membership.

Pathways represent a group of biochemical entities (e.g., genes, proteins, etc.), connected by interactions, relations, and reactions (including physical interactions, complex formation, transcriptional regulation, etc.), that lead to a certain product or changes in a cell. Molecular pathways have long been known to play a crucial role in cancer initiation, progression, dissemination, and therapeutic response. Some notable examples are: the role of RAS and PI3K pathways in prostate and breast cancers and their therapeutic responses (Yue et al., 2002; Haagenson and Wu, 2010), the Wnt signaling pathway in colorectal and other cancers (Zhan et al., 2017), the Hippo

pathway in melanoma (Zhang X. et al., 2020), and the MYC pathway in prostate cancer progression and treatment response (Arriaga et al., 2020).

Both pathCHEMO and pathER assume that interrogation of molecular pathways, such as those present in Biocarta (Nishimura, 2001), KEGG (Kanehisa et al., 2021), and Reactome (Jassal et al., 2020), can reveal functional, biologically meaningful biomarkers that govern carcinogenesis and therapeutic response. pathCHEMO was specifically developed to compare poor versus good therapeutic response (as categorical outcomes) in cancer. In general, it evaluates differential behavior of biological pathways on both transcriptomic (RNA expression) and epigenomic (DNA methylation) levels between any two phenotypes of interest (Epsi et al., 2019). First, an RNA expression treatment response signature is defined as a list of genes ranked by their differential expression between poor and good treatment response. Then, genes in each pathway are evaluated for their enrichment in either over-expressed, under-expressed, or differentially expressed (which includes both over- and under-expressed) part of this signature. Enrichment in the over- and under-expressed parts separately allows identification of pathways where the majority of genes exhibit a similar behavior (i.e., are either over- or under-expressed), while enrichment in the differentially expressed part of the signature allows identification of pathways where some genes are over-expressed and some are under-expressed (which depicts a complex interplay of activation

and repression relationships inside a molecular pathway). This enrichment is referred to as the RNA expression-based activity level of a molecular pathway. DNA methylation-based activity for each pathway is estimated in the same manner using a DNA methylation treatment response signature. Pathways that are enriched in the RNA expression treatment response signature and the DNA methylation treatment response signature are then integrated to select those that are significantly affected on both expression and methylation levels (Figure 6). Activity levels of the candidate pathways are further evaluated as biomarkers of therapeutic response in independent patient cohorts. Epsi et al. showed that pathCHEMO could successfully identify molecular pathways as biomarkers of response to commonly used chemotherapy in lung adenocarcinoma, lung squamous carcinoma, and colorectal adenocarcinoma (Epsi et al., 2019). Yet, a large number of genes that participate in these pathways could potentially preclude their adoption to clinic. To overcome this limitation, “read-out” genes for each pathway were identified for which expression levels (i) correlate with pathway activity and (ii) are associated with therapeutic response. Such read-out genes were shown to produce the same predictive accuracy as the pathways themselves and constitute feasible biomarkers for clinical use (Epsi et al., 2019). pathCHEMO is freely available at http://license.rutgers.edu/technologies/2019-121_pathchemo.

As opposed to pathCHEMO, pathER applies a pathway-based approach on a single-patient level, which allows the association of pathway activity across a patient cohort to a wide range of therapeutic responses (Rahem et al., 2020). Specifically, this approach utilizes a multivariable regression Cox proportional hazards model to associate pathway activity levels with time-to-therapeutic failure, thus capturing poor, good, and medium therapeutic responses. Rahem et al. successfully applied this approach to identify both pathways and their read-out genes for tamoxifen resistance in ER-positive breast cancer (Rahem et al., 2020). pathCHEMO and pathER were compared to other approaches, including black-box machine learning techniques (such as random forest and support vector machines) and differential gene expression alone, and were shown to outperform these approaches in identifying more accurate biomarkers of therapeutic response (Epsi et al., 2019; Rahem et al., 2020).

CHALLENGES AND LIMITATIONS OF MECHANISM-CENTRIC APPROACHES

Mechanism-centric approaches provide a powerful solution for informed biomarker discovery, yet common challenges that these methods need to account for include sufficient cohort sizes, data variability and scaling, comprehension of existing knowledge bases, and tissue-specificity (Table 1).

As many of these methods utilize association-based analyses (i.e., correlation, mutual information, regression, etc.), a sufficient cohort size is required to be able to accurately estimate relationships between variables. One of the direct solutions to this problem includes combining analyses in multiple datasets; however, batch effects among different acquisition methods,

profiling platforms, and even institutions where datasets were collected might hamper such implementation.

In addition to a sufficient cohort size, substantial variability of expression profiles is also required to be able to accurately predict associations between variables. This task is feasible, yet it requires careful consideration, meticulous initial experimental design, and in-depth investigation of the amount of final variability necessary for successful analysis. Another challenge is the need for well-defined phenotypes, as they often require a substantially large number of samples inside each phenotype group while also demanding intra-sample homogeneity, as in the eigengene approach, MARINa, PPI network-based method by Chuang et al., pathCHEMO, etc.

At the same time, methods that rely on single-patient/sample mining (e.g., VIPER, the PPI network-based method by Chuang et al., and pathER) rely on dataset scaling to define its single-sample signatures (defined by comparing each gene to the average of its expression in the dataset of interest) making interpretation of any findings from such analyses dataset-specific.

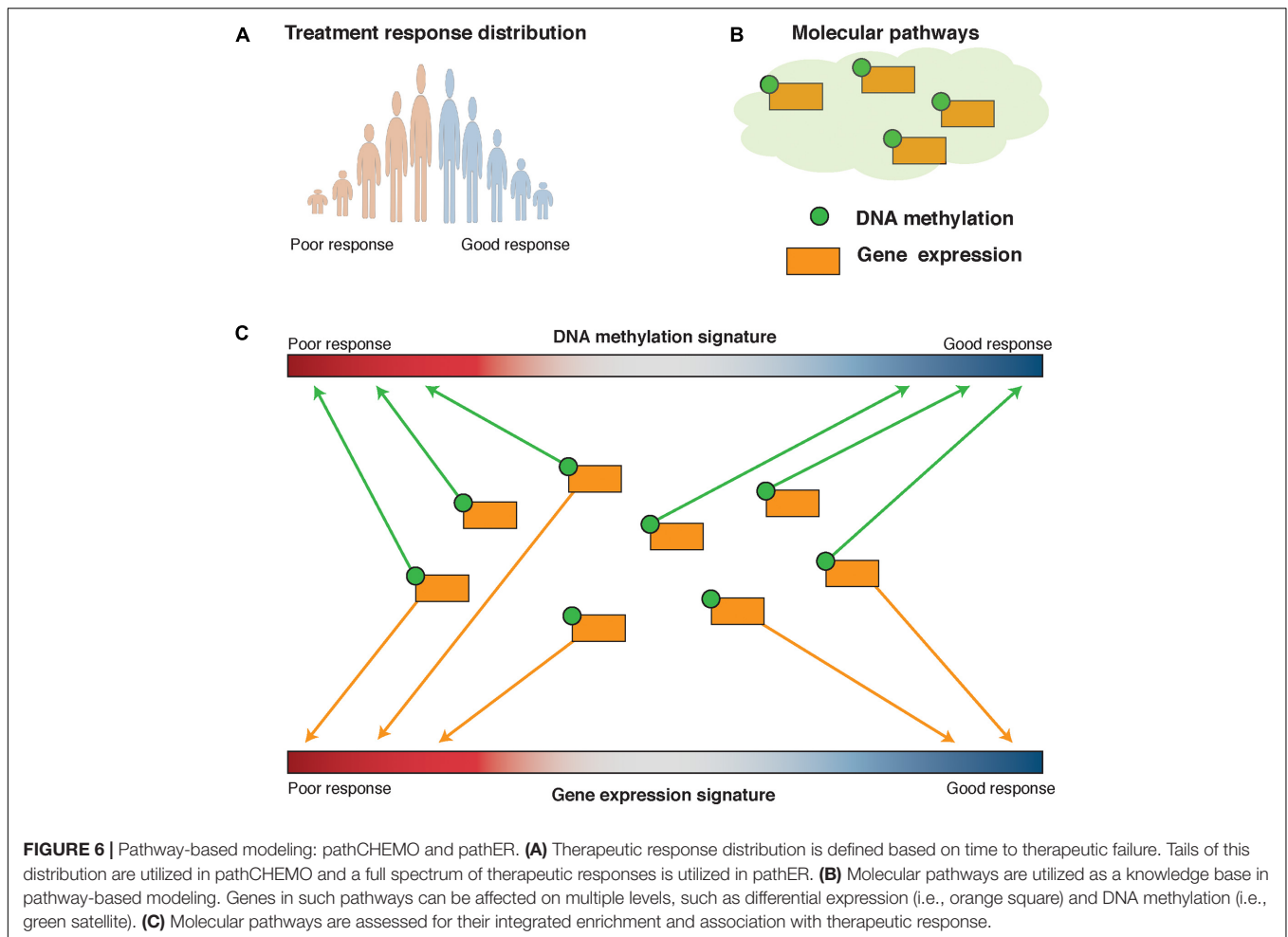
Another known challenge is tissue-specificity, commonly faced in PPI network-based and pathway-based approaches, though some tissue- and cell-specific interaction databases are now available such as TissueNet (Basha et al., 2017), the Integrated Interactions Database (Kotlyar et al., 2019), and HumanBase (Greene et al., 2015). Tissue-specificity in these methods is usually achieved by overlaying gene expression data onto the PPI networks or molecular pathways, such as in Chuang et al., pathCHEMO, and pathER.

Furthermore, limitations of mechanism-centric approaches that utilize knowledge bases (e.g., RegNetDriver, PPI network-based approach, pathCHEMO, and pathER) lie in their reliance on known biological relationships among groups of genes/proteins/other functional units contained within a database. Various annotation, pathway, and PPI databases depend on existing information and do not include functional units that have not been previously studied, thus limiting *de novo* discoveries.

DISCUSSION

The wide availability of large-scale data produced by high-throughput technologies has created a wealth of information for biomarker discovery. A vast majority of these biomarkers have been identified using gene-centric methods, yet their interpretability and clinical utility have been limited as they do not account for the relationships among genes. Utilizing methods that consider biological underpinnings of the data (i.e., mechanism-centric methods) can vastly improve interpretable biomarker discovery, clinical applicability and targeting, and reproducibility of results.

In particular, advantages of mechanism-centric over gene-centric approaches can be illustrated through their ability to (i) identify a tightly connected, cooperative group of genes unified by the same function, as opposed to individual genes (which might not be related); (ii) provide a mechanism-level view, which enhances the understanding of the biological mechanisms



implicated in a phenotype (e.g., therapeutic resistance, cancer metastasis); (iii) look at alterations in biological structures, which enhances the likelihood of identifying functionally relevant targets; (iv) identify driver as opposed to passenger markers, which allows for their effective therapeutic targeting; (v) focus on molecular structures, rather than individual genes, which decreases the chance of detecting results due to experimental noise present in biological experiments (i.e., robustness of results); and finally (vi) identify biomarkers that are more accurate and more reproducible between different cohorts.

From a computational point of view, mechanism-centric approaches can be used for interpretable feature engineering and selection (i.e., reduction), subsequently reducing the number of hypotheses to be tested. This is clearly demonstrated by gene co-expression networks, regulatory networks, PPI networks, and pathway-based methods, where cooperative groups of genes, instead of a long list of singular genes, are assessed for their association with clinical outcomes.

Mechanism-centric methods can both (i) provide interpretable inputs to white- or black-box approaches or (ii) contribute to inner model interpretability (i.e., such as in visible machine learning). First, results from mechanism-centric

methods can be utilized as inputs into learning models to significantly improve predictive performance (over gene-centric inputs). One such example was demonstrated in Rahem et al., where pathway-based markers were utilized as inputs into Cox proportional hazards regression modeling and outperformed gene-centric markers for tamoxifen resistance in ER-positive breast cancer (Rahem et al., 2020). Similarly, Chuang et al. showed that markers identified by their PPI network-based method could be effectively used as inputs into a regression model and outperformed gene-centric markers in classification of metastatic breast cancer (Chuang et al., 2007). Though not in cancer, several methods have also suggested utilizing hierarchical structures (such as those inherent in Gene Ontology) as inputs for predictive models (Carvunis and Ideker, 2014; Yu et al., 2016). Second, mechanism-centric methods can potentially be incorporated into model building, such as in “visible learning,” where the relationships between inputs and outputs can be interpreted (Yu et al., 2018). One such (outside of cancer) neural network method, DCell, was proposed by Ma et al., where the hierarchy of molecular relationships determined from prior knowledge (Gene Ontology and CliXO) was built into the model itself (i.e., hierarchies were utilized by nodes of the neural

TABLE 1 | Summary of mechanism-centric methods discussed in this review.

Method	Data modality	Utilize knowledge base?
Gene co-expression network-based	<i>Identify modules of highly correlated genes</i> +Increased interpretability at the mechanistic level +Associate genes with previously uncharacterized biological functions -Directionality of gene-gene interactions is unknown	
Centered Concordance Index (CCI) (Han et al., 2016)	<i>Condition-specific module identification</i> Single-omic	No
Eigengenes (Alter et al., 2000; Zhang and Horvath, 2005)	<i>Identify modules associated with clinical features of interest</i> Single-omic	No
Hubs (Freeman, 1977; Horvath and Dong, 2008)	<i>Hub gene identification</i> +Identify potential mechanism-centric target Single-omic	No
Regulatory network-based	<i>Identify regulatory relationships between a TF/co-TF and its target genes</i> +Increased interpretability at the mechanistic level +Identify potential drivers of disease +Can identify non-linear relationships +Tissue specific network	
MARINa (Lefebvre et al., 2010)	<i>Identify MRs from a set of samples containing two phenotypes</i> -Need phenotype signature Single-omic	No
VIPER (Alvarez et al., 2016)	<i>Single-sample MR identification from a cohort</i> -Dataset scaling Single-omic	No
RegNetDriver (Dhingra et al., 2017)	<i>Identify TF hubs that are significantly affected by single nucleotide variants, structural variants, or DNA methylation</i> +Increase interpretability of TF hub activity through multi-omic integration -Limited by information in knowledge base Multi-omic	Yes
PPI network-based	<i>Use PPI subnetworks as a functional unit</i> +Increased interpretability at the mechanistic level +Connect results to the protein complex level -Limited by information in knowledge base	
Chuang et al., 2007	<i>Identify subnetworks with differential activity in metastatic breast cancer</i> +Tissue-specificity from overlaying gene expression data +Improved biomarker classification accuracy and reproducibility -Dataset scaling Multi-omic	Yes
Pathway-based	<i>Use molecular pathways as a functional unit</i> +Increased interpretability at the mechanistic level -Limited by information in knowledge base	
pathCHEMO (Epsi et al., 2019)	<i>Identify significantly altered pathways (at transcript and DNA methylation levels) in response to chemotherapy in lung and colorectal cancer</i> +Improved biomarker classification accuracy and reproducibility -Need phenotype signature Multi-omic	Yes
pathER (Rahem et al., 2020)	<i>Identify pathways as markers of tamoxifen resistance in ER + breast cancer</i> +Improved biomarker classification accuracy and reproducibility -Dataset scaling Single-omic	Yes

The objective of each method is detailed in italics, followed by their respective pros (+) and cons (-). Overall pros and cons for each method type are listed in a non-redundant manner. Information on data modality and if a method utilized a knowledge base is detailed as well.

network) (Ma et al., 2018). Recently, Kuenzi et al. developed an extension of DCell, called DrugCell, which utilized chemical drug structures as a part of the neural network learning model to predict drug response in cancer cells (Kuenzi et al., 2020). This interpretable deep learning model was shown to be able to

predict cell sensitivity/resistance to specific drugs, synergistic drug mechanisms, and effective drug combinations for treatment.

Further improvements in the interpretability of biological processes that inform discovery of mechanism-centric biomarkers can be made through multi-level data and method

integration. For example, several groups have combined co-expression WGCNA modules with PPI networks to uncover hubs with functional connections as biomarkers in endometrial cancer (Liu et al., 2019) and bladder cancer (Wang Y. et al., 2019). Wang et al. constructed an Active Protein-Gene network model using transcriptional regulatory and PPI networks to quantify TF activity and elucidate both upstream and downstream regulations (Wang et al., 2013). Even though this study was done in diabetes, it could be applicable to mechanism-centric biomarker discovery in cancer. Ahsen et al. embedded VIPER within a new framework (NeTFactor) to identify TFs that most likely regulate a gene-centric biomarker signature (Ahsen et al., 2019). While this method was applied to asthma and peanut allergy, it could easily be extended to cancer studies. At the same time, multi-omic integration in RegNetDriver improved the interpretability of the proposed model to explain the impact of mutations, structural variants, and DNA methylation on TF activity in prostate cancer (Dhingra et al., 2017). A recent study by Broyde et al. constructed a multi-omic lung adenocarcinoma tissue-specific oncoprotein interaction network using information obtained from ARACNe, CINDy (an algorithm identifying post-translational modulators), VIPER, and PPI predictions (Broyde et al., 2021), which depicted a complex network of interactions for KRAS and could potentially be utilized for mechanism-centric biomarker discovery. Such multi-level approaches in conjunction with mechanism-centric methods promise to uncover a deeper understanding of mechanisms involved in gene regulation and post-translational modifications in biomarker discovery.

Finally, recent technological advances, such as those seen in single-cell studies, promise to improve our understanding of intra-tumor heterogeneity, clonal evolution, and the role of microenvironment in cancer progression and therapeutic response. Single-cell gene expression offers a granular view of active pathways in a cell type-specific manner and potentially allows for the construction of cell type-specific networks. In fact, the rapid advances of single-cell sequencing technology have already allowed network analysis methods to be applied directly to data from single-cell RNA-sequencing (scRNA-seq) (Crow et al., 2016; Aibar et al., 2017; Chan et al., 2017; Fiers et al., 2018; Papili Gao et al., 2018; van Dijk et al., 2018; Lamere and Li, 2019; Jackson et al., 2020;

Sekula et al., 2020; Ye et al., 2020) with integration of other data modalities for improved network inference (Aibar et al., 2017; Chan et al., 2017; Papili Gao et al., 2018; van Dijk et al., 2018; Jackson et al., 2020; Pratapa et al., 2020). Furthermore, matching single-cell and bulk patient samples could provide an invaluable resource for single-cell driven network investigations that can be compared to and related back to bulk tissues. As more single-cell data become available (e.g., RNA sequencing, targeted DNA sequencing, ATAC-seq, etc.), we foresee advances in single-cell technologies and data analysis to be central to understanding precise, clone-specific biomarkers, unveiling trajectories of tumor evolution and providing accurate ground for informed time-cautious precision therapeutics.

In summary, mechanism-centric approaches (based on gene co-expression networks, regulatory networks, PPI networks, and molecular pathways) identify biomarkers that are biologically meaningful, interpretable, reproducible, have higher translational potential, and provide greater predictive power over biomarkers identified by gene-centric methods. Thus, mechanism-centric approaches are the future of clinically relevant rational biomarker discovery, personalized treatment planning, and precision therapeutics in cancer.

AUTHOR CONTRIBUTIONS

CY and AM conceived and wrote the manuscript. Both the authors contributed to the article and approved the submitted version.

FUNDING

AM was supported by R01LM013236-01 and Rutgers start-up funds.

ACKNOWLEDGMENTS

We are thankful to the Mitrofanova lab for useful discussions.

REFERENCES

- Abida, W., Cyrta, J., Heller, G., Prandi, D., Armenia, J., Coleman, I., et al. (2019). Genomic correlates of clinical outcome in advanced prostate cancer. *Proc. Natl. Acad. Sci. U.S.A.* 116, 11428–11436.
- Agnelli, L., Forcato, M., Ferrari, F., Tuana, G., Todoerti, K., Walker, B. A., et al. (2011). The reconstruction of transcriptional networks reveals critical genes with implications for clinical outcome of multiple myeloma. *Clin. Cancer Res.* 17, 7402–7412. doi: 10.1158/1078-0432.ccr-11-0596
- Ahsen, M. E., Chun, Y., Grishin, A., Grishina, G., Stolovitzky, G., Pandey, G., et al. (2019). NeTFactor, a framework for identifying transcriptional regulators of gene expression-based biomarkers. *Sci. Rep.* 9:12970.
- Aibar, S., Gonzalez-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulsemans, G., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* 14, 1083–1086. doi: 10.1038/nmeth.4463
- Allen, W. L., Coyle, V. M., and Johnston, P. G. (2006). Predicting the outcome of chemotherapy for colorectal cancer. *Curr. Opin. Pharmacol.* 6, 332–336. doi: 10.1016/j.coph.2006.02.005
- Alter, O., Brown, P. O., and Botstein, D. (2000). Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl. Acad. Sci. U.S.A.* 97, 10101–10106. doi: 10.1073/pnas.97.18.10101
- Alvarez, M. J., Shen, Y., Giorgi, F. M., Lachmann, A., Ding, B. B., Ye, B. H., et al. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* 48, 838–847. doi: 10.1038/ng.3593
- Arriaga, J. M., Panja, S., Alshalalfa, M., Zhao, J., Zou, M., Giacobbe, A., et al. (2020). A MYC and RAS co-activation signature in localized prostate cancer drives bone metastasis and castration resistance. *Nat. Cancer* 1, 1082–1096. doi: 10.1038/s43018-020-00125-0
- Ayer, T., Alagoz, O., Chhatwal, J., Shavlik, J. W., Kahn, C. E. Jr., and Burnside, E. S. (2010). Breast cancer risk estimation with artificial neural networks revisited:

- discrimination and calibration. *Cancer* 116, 3310–3321. doi: 10.1002/cncr.25081
- Ayers, M., Symmans, W. F., Stec, J., Damokosh, A. I., Clark, E., Hess, K., et al. (2004). Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. *J. Clin. Oncol.* 22, 2284–2293. doi: 10.1200/jco.2004.05.166
- Aytes, A., Giacobbe, A., Mitrofanova, A., Ruggiero, K., Cyrt, J., Arriaga, J., et al. (2018). NSD2 is a conserved driver of metastatic prostate cancer progression. *Nat. Commun.* 9:5201.
- Aytes, A., Mitrofanova, A., Lefebvre, C., Alvarez, M. J., Castillo-Martin, M., Zheng, T., et al. (2014). Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell* 25, 638–651. doi: 10.1016/j.ccr.2014.03.017
- Bae, T., Rho, K., Choi, J. W., Horimoto, K., Kim, W., and Kim, S. (2013). Identification of upstream regulators for prognostic expression signature genes in colorectal cancer. *BMC Syst. Biol.* 7:86. doi: 10.1186/1752-0509-7-86
- Basha, O., Barshir, R., Sharon, M., Lerman, E., Kirson, B. F., Hekselman, I., et al. (2017). The TissueNet v.2 database: a quantitative view of protein-protein interactions across human tissues. *Nucleic Acids Res.* 45, D427–D431.
- Basso, K., Margolin, A. A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. *Nat. Genet.* 37, 382–390. doi: 10.1038/ng1532
- Beer, D. G., Kardia, S. L., Huang, C. C., Giordano, T. J., Levin, A. M., Misek, D. E., et al. (2002). Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat. Med.* 8, 816–824.
- Bisikirska, B., Bansal, M., Shen, Y., Teruya-Feldstein, J., Chaganti, R., and Califano, A. (2016). Elucidation and pharmacological targeting of novel molecular drivers of follicular lymphoma progression. *Cancer Res.* 76, 664–674. doi: 10.1158/0008-5472.can-15-0828
- Boumahdi, S., Driessens, G., Lapouge, G., Rorive, S., Nassar, D., Le Mercier, M., et al. (2014). SOX2 controls tumour initiation and cancer stem-cell functions in squamous-cell carcinoma. *Nature* 511, 246–250. doi: 10.1038/nature13305
- Boutros, P. C. (2015). The path to routine use of genomic biomarkers in the cancer clinic. *Genome Res.* 25, 1508–1513. doi: 10.1101/gr.191114.115
- Broyde, J., Simpson, D. R., Murray, D., Paull, E. O., Chu, B. W., Tagore, S., et al. (2021). Oncoprotein-specific molecular interaction maps (SigMaps) for cancer network analyses. *Nat. Biotechnol.* 39, 215–224. doi: 10.1038/s41587-020-0652-7
- Butte, A. J., and Kohane, I. S. (2000). Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pac. Symp. Biocomput.* 5, 418–429.
- Butte, A. J., Tamayo, P., Slonim, D., Golub, T. R., and Kohane, I. S. (2000). Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc. Natl. Acad. Sci. U.S.A.* 97, 12182–12186. doi: 10.1073/pnas.220392197
- Carro, M. S., Lim, W. K., Alvarez, M. J., Bollo, R. J., Zhao, X., Snyder, E. Y., et al. (2010). The transcriptional network for mesenchymal transformation of brain tumours. *Nature* 463, 318–325. doi: 10.1038/nature08712
- Carvunis, A. R., and Ideker, T. (2014). Siri of the cell: what biology could learn from the iPhone. *Cell* 157, 534–538. doi: 10.1016/j.cell.2014.03.009
- Chan, T. E., Stumpf, M. P. H., and Babbie, A. C. (2017). Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Syst.* 5, 251–267.e3.
- Chen, Y. I., Moore, R. E., Ge, H. Y., Young, M. K., Lee, T. D., and Stevens, S. W. (2007). Proteomic analysis of in vivo-assembled pre-mRNA splicing complexes expands the catalog of participating factors. *Nucleic Acids Res.* 35, 3928–3944. doi: 10.1093/nar/gkm347
- Chng, W. J., Chung, T. H., Kumar, S., Usmani, S., Munshi, N., Avet-Loiseau, H., et al. (2016). Gene signature combinations improve prognostic stratification of multiple myeloma patients. *Leukemia* 30, 1071–1078. doi: 10.1038/leu.2015.341
- Chuang, H. Y., Lee, E., Liu, Y. T., Lee, D., and Ideker, T. (2007). Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.* 3:140. doi: 10.1038/msb4100180
- Cordero, D., Sole, X., Crous-Bou, M., Sanz-Pamplona, R., Pare-Brunet, L., Guino, E., et al. (2014). Large differences in global transcriptional regulatory programs of normal and tumor colon cells. *BMC Cancer* 14:708. doi: 10.1186/1471-2407-14-708
- Cowley, M. J., Pinese, M., Kassahn, K. S., Waddell, N., Pearson, J. V., Grimmond, S. M., et al. (2012). PINA v2.0: mining interactome modules. *Nucleic Acids Res.* 40, D862–D865.
- Crow, M., Paul, A., Ballouz, S., Huang, Z. J., and Gillis, J. (2016). Exploiting single-cell expression to characterize co-expression replicability. *Genome Biol.* 17:101.
- Della Gatta, G., Palomero, T., Perez-Garcia, A., Ambesi-Impiombato, A., Bansal, M., Carpenter, Z. W., et al. (2012). Reverse engineering of TLX oncogenic transcriptional networks identifies RUNX1 as tumor suppressor in T-ALL. *Nat. Med.* 18, 436–440. doi: 10.1038/nm.2610
- Dhingra, P., Martinez-Fundichely, A., Berger, A., Huang, F. W., Forbes, A. N., Liu, E. M., et al. (2017). Identification of novel prostate cancer drivers using RegNetDriver: a framework for integration of genetic and epigenetic alterations with tissue-specific regulatory network. *Genome Biol.* 18:141.
- Emmert-Streib, F., Dehmer, M., and Haibe-Kains, B. (2014). Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Front. Cell Dev. Biol.* 2:38. doi: 10.3389/fcell.2014.00038
- Encode Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. doi: 10.1038/nature11247
- Epsi, N. J., Panja, S., Pine, S. R., and Mitrofanova, A. (2019). pathCHEMO, a generalizable computational framework uncovers molecular pathways of chemoresistance in lung adenocarcinoma. *Commun. Biol.* 2:334.
- Erho, N., Crisan, A., Vergara, I. A., Mitra, A. P., Ghadessi, M., Buerki, C., et al. (2013). Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. *PLoS One* 8:e66855. doi: 10.1371/journal.pone.0066855
- Eskandari, E., Mahjoubi, F., and Motalebzadeh, J. (2018). An integrated study on TFs and miRNAs in colorectal cancer metastasis and evaluation of three co-regulated candidate genes as prognostic markers. *Gene* 679, 150–159. doi: 10.1016/j.gene.2018.09.003
- Fiers, M., Minnoye, L., Aibar, S., Bravo Gonzalez-Blas, C., Kalender Atak, Z., and Aerts, S. (2018). Mapping gene regulatory networks from single-cell omics data. *Brief. Funct. Genomics* 17, 246–254. doi: 10.1093/bfpg/elx046
- Fletcher, M. N., Castro, M. A., Wang, X., de Santiago, I., O'Reilly, M., Chin, S. F., et al. (2013). Master regulators of FGFR2 signalling and breast cancer risk. *Nat. Commun.* 4:2464.
- Freeman, L. C. A. (1977). Set of measures of centrality based on betweenness. *Sociometry* 40, 35–41. doi: 10.2307/3033543
- Friedman, N., Linial, M., Nachman, I., and Pe'er, D. (2000). Using bayesian networks to analyze expression data. *J. Comput. Biol.* 7, 601–620. doi: 10.1089/106652700750050961
- Gabay, M., Li, Y., and Felsher, D. W. (2014). MYC activation is a hallmark of cancer initiation and maintenance. *Cold Spring Harb. Perspect. Med.* 4:a014241. doi: 10.1101/cshperspect.a014241
- Garzotto, M., Beer, T. M., Hudson, R. G., Peters, L., Hsieh, Y. C., Barrera, E., et al. (2005). Improved detection of prostate cancer using classification and regression tree analysis. *J. Clin. Oncol.* 23, 4322–4329. doi: 10.1200/jco.2005.11.136
- Giulietti, M., Occhipinti, G., Principato, G., and Piva, F. (2016). Weighted gene co-expression network analysis reveals key genes involved in pancreatic ductal adenocarcinoma development. *Cell Oncol.* 39, 379–388. doi: 10.1007/s13402-016-0283-7
- Giulietti, M., Occhipinti, G., Principato, G., and Piva, F. (2017). Identification of candidate miRNA biomarkers for pancreatic ductal adenocarcinoma by weighted gene co-expression network analysis. *Cell Oncol.* 40, 181–192. doi: 10.1007/s13402-017-0315-y
- Greber, B. J., and Nogales, E. (2019). The structures of eukaryotic transcription pre-initiation complexes and their functional implications. *Subcell. Biochem.* 93, 143–192. doi: 10.1007/978-3-030-28151-9_5
- Greene, C. S., Krishnan, A., Wong, A. K., Ricciotti, E., Zelaya, R. A., Himmelstein, D. S., et al. (2015). Understanding multicellular function and disease with human tissue-specific networks. *Nat. Genet.* 47, 569–576. doi: 10.1038/ng.3259
- Haagensen, K. K., and Wu, G. S. (2010). The role of MAP kinases and MAP kinase phosphatase-1 in resistance to breast cancer treatment. *Cancer Metastasis Rev.* 29, 143–149. doi: 10.1007/s10555-010-9208-5

- Han, Y., Ye, X., Cheng, J., Zhang, S., Feng, W., Han, Z., et al. (2019). Integrative analysis based on survival associated co-expression gene modules for predicting Neuroblastoma patients' survival time. *Biol. Direct* 14:4.
- Han, Z., Zhang, J., Sun, G., Liu, G., and Huang, K. (2016). A matrix rank based concordance index for evaluating and detecting conditional specific co-expressed gene modules. *BMC Genomics* 17(Suppl. 7):519. doi: 10.1186/s12864-016-2912-y
- Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kocicinski, F., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 22, 1760–1774. doi: 10.1101/gr.135350.111
- Hecker, M., Lambeck, S., Toepfer, S., van Someren, E., and Guthke, R. (2009). Gene regulatory network inference: data integration in dynamic models—a review. *Biosystems* 96, 86–103.
- Heng, Y. J., Lester, S. C., Tse, G. M., Factor, R. E., Allison, K. H., Collins, L. C., et al. (2017). The molecular basis of breast cancer pathological phenotypes. *J. Pathol.* 241, 375–391. doi: 10.1002/path.4847
- Hoadley, K. A., Yau, C., Hinoue, T., Wolf, D. M., Lazar, A. J., Drill, E., et al. (2018). Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* 173, 291–304.e6.
- Horvath, S., and Dong, J. (2008). Geometric interpretation of gene coexpression network analysis. *PLoS Comput. Biol.* 4:e1000117. doi: 10.1371/journal.pcbi.1000117
- Horvath, S., Zhang, B., Carlson, M., Lu, K. V., Zhu, S., Felciano, R. M., et al. (2006). Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc. Natl. Acad. Sci. U.S.A.* 103, 17402–17407. doi: 10.1073/pnas.0608396103
- Hu, X., Bao, M., Huang, J., Zhou, L., and Zheng, S. (2020). Identification and validation of novel biomarkers for diagnosis and prognosis of hepatocellular carcinoma. *Front. Oncol.* 10:541479. doi: 10.3389/fonc.2020.541479
- Huang, J. K., Carlin, D. E., Yu, M. K., Zhang, W., Kreisberg, J. F., Tamayo, P., et al. (2018). Systematic evaluation of molecular networks for discovery of disease genes. *Cell Syst.* 6, 484–495.e5.
- Huang, Z., Han, Z., Wang Resource, T., Shao, W., Xiang, S., Salama, P., et al. (2021). TSUNAMI: translational bioinformatics tool suite for network analysis and mining. *Genomics Proteomics Bioinformatics.*
- Huo, T., Canepa, R., Sura, A., Modave, F., and Gong, Y. (2017). Colorectal cancer stages transcriptome analysis. *PLoS One* 12:e0188697. doi: 10.1371/journal.pone.0188697
- Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. *PLoS One* 5:e12776. doi: 10.1371/journal.pone.0012776
- Jackson, C. A., Castro, D. M., Saldi, G. A., Bonneau, R., and Gresham, D. (2020). Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments. *Elife* 9:e51254.
- Jain, R. K., Duda, D. G., Willett, C. G., Sahani, D. V., Zhu, A. X., Loeffler, J. S., et al. (2009). Biomarkers of response and resistance to antiangiogenic therapy. *Nat. Rev. Clin. Oncol.* 6, 327–338.
- Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., et al. (2020). The reactome pathway knowledgebase. *Nucleic Acids Res.* 48, D498–D503.
- Jeong, H., Mason, S. P., Barabasi, A. L., and Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature* 411, 41–42. doi: 10.1038/35075138
- Jia, R., Zhao, H., and Jia, M. (2020). Identification of co-expression modules and potential biomarkers of breast cancer by WGCNA. *Gene* 750:144757. doi: 10.1016/j.gene.2020.144757
- Jiramongkol, Y., and Lam, E. W. (2020). FOXO transcription factor family in cancer and metastasis. *Cancer Metastasis Rev.* 39, 681–709. doi: 10.1007/s10555-020-09883-w
- Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 49, D545–D551.
- Karlebach, G., and Shamir, R. (2008). Modelling and analysis of gene regulatory networks. *Nat. Rev. Mol. Cell Biol.* 9, 770–780.
- Kotlyar, M., Pastrello, C., Malik, Z., and Jurisica, I. (2019). IID 2018 update: context-specific physical protein-protein interactions in human, model organisms and domesticated species. *Nucleic Acids Res.* 47, D581–D589.
- Kuenzi, B. M., Park, J., Fong, S. H., Sanchez, K. S., Lee, J., Kreisberg, J. F., et al. (2020). Predicting drug response and synergy using a deep learning model of human cancer cells. *Cancer Cell* 38, 672–684.e6.
- Kuiper, R., Broyl, A., de Knecht, Y., van Vliet, M. H., van Beers, E. H., van der Holt, B., et al. (2012). A gene expression signature for high-risk multiple myeloma. *Leukemia* 26, 2406–2413.
- Lamere, A. T., and Li, J. (2019). Inference of gene co-expression networks from single-cell RNA-sequencing data. *Methods Mol. Biol.* 1935, 141–153. doi: 10.1007/978-1-4939-9057-3_10
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559
- Lee, W. P., and Tzou, W. S. (2009). Computational methods for discovering gene networks from expression data. *Brief. Bioinform.* 10, 408–423.
- Lefebvre, C., Rajbhandari, P., Alvarez, M. J., Bandaru, P., Lim, W. K., Sato, M., et al. (2010). A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol. Syst. Biol.* 6:377. doi: 10.1038/msb.2010.31
- Lim, W. K., Lyashenko, E., and Califano, A. (2009). Master regulators used as breast cancer metastasis classifier. *Pac. Symp. Biocomput.* 14, 504–515.
- Liu, J., Zhou, S., Li, S., Jiang, Y., Wan, Y., Ma, X., et al. (2019). Eleven genes associated with progression and prognosis of endometrial cancer (EC) identified by comprehensive bioinformatics analysis. *Cancer Cell Int.* 19, 136.
- Liu, R., Guo, C. X., and Zhou, H. H. (2015a). Network-based approach to identify prognostic biomarkers for estrogen receptor-positive breast cancer treatment with tamoxifen. *Cancer Biol. Ther.* 16, 317–324. doi: 10.1080/15384047.2014.1002360
- Liu, R., Lv, Q. L., Yu, J., Hu, L., Zhang, L. H., Cheng, Y., et al. (2015b). Correlating transcriptional networks with pathological complete response following neoadjuvant chemotherapy for breast cancer. *Breast Cancer Res. Treat.* 151, 607–618. doi: 10.1007/s10549-015-3428-x
- Ludwig, J. A., and Weinstein, J. N. (2005). Biomarkers in cancer staging, prognosis and treatment selection. *Nat. Rev. Cancer* 5, 845–856. doi: 10.1038/nrc1739
- Ma, J., Yu, M. K., Fong, S., Ono, K., Sage, E., Demchak, B., et al. (2018). Using deep learning to model the hierarchical structure and function of a cell. *Nat. Methods* 15, 290–298. doi: 10.1038/nmeth.4627
- Mak, H. C., Daly, M., Gruebel, B., and Ideker, T. (2007). CellCircuits: a database of protein network models. *Nucleic Acids Res.* 35, D538–D545.
- Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., et al. (2006a). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7(Suppl. 1):S7. doi: 10.1186/1471-2105-7-S1-S7
- Margolin, A. A., Wang, K., Lim, W. K., Kustagi, M., Nemenman, I., and Califano, A. (2006b). Reverse engineering cellular networks. *Nat. Protoc.* 1, 662–671.
- Markowitz, F., and Spang, R. (2007). Inferring cellular networks—a review. *BMC Bioinformatics* 8(Suppl. 6):S5. doi: 10.1186/1471-2105-8-S6-S5
- McDermott, J. E., Wang, J., Mitchell, H., Webb-Robertson, B. J., Hafen, R., Ramey, J., et al. (2013). Challenges in biomarker discovery: combining expert insights with statistical analysis of complex omics data. *Expert Opin. Med. Diagn.* 7, 37–51. doi: 10.1517/17530059.2012.718329
- Michiels, S., Koscielny, S., and Hill, C. (2005). Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 365, 488–492. doi: 10.1016/s0140-6736(05)17866-0
- Miryala, S. K., Anbarasu, A., and Ramaiah, S. (2018). Discerning molecular interactions: a comprehensive review on biomolecular interaction databases and network analysis tools. *Gene* 642, 84–94. doi: 10.1016/j.gene.2017.11.028
- Mitrofanova, A., Aytes, A., Zou, M., Shen, M. M., Abate-Shen, C., and Califano, A. (2015). Predicting drug response in human prostate cancer from preclinical analysis of in vivo mouse models. *Cell Rep.* 12, 2060–2071. doi: 10.1016/j.celrep.2015.08.051
- Nishimura, D. (2001). BioCarta. *Biotech. Softw. Internet Rep.* 2, 117–120. doi: 10.1089/152791601750294344
- Orchard, S., Amari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., et al. (2014). The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* 42, D358–D363.
- Ou, Y., and Zhang, C.-Q. (2007). A new multimembership clustering method. *J. Ind. Manage. Optim.* 3, 619–624. doi: 10.3934/jimo.2007.3.619
- Palomero, T., Lim, W. K., Odom, D. T., Sulis, M. L., Real, P. J., Margolin, A., et al. (2006). NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. *Proc. Natl. Acad. Sci. U.S.A.* 103, 18261–18266. doi: 10.1073/pnas.0606108103

- Panja, S., Hayati, S., Epsi, N. J., Parrott, J. S., and Mitrofanova, A. (2018). Integrative (epi) genomic analysis to predict response to androgen-deprivation therapy in prostate cancer. *EBioMedicine* 31, 110–121. doi: 10.1016/j.ebiom.2018.04.007
- Panja, S., Rahem, S., Chu, C. J., and Mitrofanova, A. (2020). Big data to knowledge: application of machine learning to predictive modeling of therapeutic response in cancer. *Curr. Genomics* 21, 1–25. doi: 10.1201/b11508-2
- Papili Gao, N., Ud-Dean, S. M. M., Gandrillon, O., and Gunawan, R. (2018). SINCERITIES: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles. *Bioinformatics* 34, 258–266. doi: 10.1093/bioinformatics/btx575
- Petty, R. D., Samuel, L. M., Murray, G. I., MacDonald, G., O’Kelly, T., Loudon, M., et al. (2009). APRIL is a novel clinical chemo-resistance biomarker in colorectal adenocarcinoma identified by gene expression profiling. *BMC Cancer* 9:434. doi: 10.1186/1471-2407-9-434
- Pratapa, A., Jalihal, A. P., Law, J. N., Bharadwaj, A., and Murali, T. M. (2020). Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods* 17, 147–154. doi: 10.1038/s41592-019-0690-6
- Rahem, S. M., Epsi, N. J., Coffman, F. D., and Mitrofanova, A. (2020). Genome-wide analysis of therapeutic response uncovers molecular pathways governing tamoxifen resistance in ER+ breast cancer. *EBioMedicine* 61:103047. doi: 10.1016/j.ebiom.2020.103047
- Remo, A., Simeone, I., Pancione, M., Parcesepo, P., Finetti, P., Cerulo, L., et al. (2015). Systems biology analysis reveals NFAT5 as a novel biomarker and master regulator of inflammatory breast cancer. *J. Transl. Med.* 13:138.
- Robichaud, N., Sonenberg, N., Ruggero, D., and Schneider, R. J. (2019). Translational control in cancer. *Cold Spring Harb. Perspect. Biol.* 11:a032896.
- Robinson, D., Van Allen, E. M., Wu, Y. M., Schultz, N., Lonigro, R. J., Mosquera, J. M., et al. (2015). Integrative clinical genomics of advanced prostate cancer. *Cell* 161, 1215–1228.
- Rosenfeld, N., Aharonov, R., Meiri, E., Rosenwald, S., Spector, Y., Zepeniuk, M., et al. (2008). MicroRNAs accurately identify cancer tissue origin. *Nat. Biotechnol.* 26, 462–469.
- Sanz-Pamplona, R., Berenguer, A., Cordero, D., Mollevi, D. G., Crous-Bou, M., Sole, X., et al. (2014). Aberrant gene expression in mucosa adjacent to tumor reveals a molecular crosstalk in colon cancer. *Mol. Cancer* 13:46. doi: 10.1186/1476-4598-13-46
- Sartor, I. T., Zeidan-Chulia, F., Albanus, R. D., Dalmolin, R. J., and Moreira, J. C. (2014). Computational analyses reveal a prognostic impact of TULP3 as a transcriptional master regulator in pancreatic ductal adenocarcinoma. *Mol. Biosyst.* 10, 1461–1468. doi: 10.1039/c3mb70590k
- Sekula, M., Gaskins, J., and Datta, S. (2020). A sparse bayesian factor model for the construction of gene co-expression networks from single-cell RNA sequencing count data. *BMC Bioinformatics* 21:361. doi: 10.1186/s12859-020-03707-y
- Shaughnessy, J. D. Jr., Qu, P., Usmani, S., Heuck, C. J., Zhang, Q., Zhou, Y., et al. (2011). Pharmacogenomics of bortezomib test-dosing identifies hyperexpression of proteasome genes, especially PSMD4, as novel high-risk feature in myeloma treated with total therapy 3. *Blood* 118, 3512–3524. doi: 10.1182/blood-2010-12-328252
- Shaughnessy, J. D. Jr., Zhan, F., Burington, B. E., Huang, Y., Colla, S., Hanamura, I., et al. (2007). A validated gene expression model of high-risk multiple myeloma is defined by deregulated expression of genes mapping to chromosome 1. *Blood* 109, 2276–2284. doi: 10.1182/blood-2006-07-038430
- Sonabend, A. M., Bansal, M., Guarnieri, P., Lei, L., Amendolara, B., Soderquist, C., et al. (2014). The transcriptional regulatory network of proneural glioma determines the genetic alterations selected during tumor progression. *Cancer Res.* 74, 1440–1451. doi: 10.1158/0008-5472.can-13-2150
- Song, H., Ding, N., Li, S., Liao, J., Xie, A., Yu, Y., et al. (2020). Identification of hub genes associated with hepatocellular carcinoma using robust rank aggregation combined with weighted gene co-expression network analysis. *Front. Genet.* 11:895. doi: 10.3389/fgene.2020.00895
- Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U.S.A.* 98, 10869–10874. doi: 10.1073/pnas.191367098
- Sotiriou, C., Neo, S. Y., McShane, L. M., Korn, E. L., Long, P. M., Jazaeri, A., et al. (2003). Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc. Natl. Acad. Sci. U.S.A.* 100, 10393–10398. doi: 10.1073/pnas.1732912100
- Strimbu, K., and Tavel, J. A. (2010). What are biomarkers? *Curr. Opin. HIV AIDS* 5, 463–466.
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613.
- Talos, F., Mitrofanova, A., Bergren, S. K., Califano, A., and Shen, M. M. (2017). A computational systems approach identifies synergistic specification genes that facilitate lineage conversion to prostate tissue. *Nat. Commun.* 8:14662.
- Tang, J., Kong, D., Cui, Q., Wang, K., Zhang, D., Gong, Y., et al. (2018). Prognostic genes of breast cancer identified by gene co-expression network analysis. *Front. Oncol.* 8:374. doi: 10.3389/fonc.2018.00374
- Tang, X., Xu, P., Wang, B., Luo, J., Fu, R., Huang, K., et al. (2019). Identification of a specific gene module for predicting prognosis in glioblastoma patients. *Front. Oncol.* 9:812. doi: 10.3389/fonc.2019.00812
- Tian, Z., He, W., Tang, J., Liao, X., Yang, Q., Wu, Y., et al. (2020). Identification of important modules and biomarkers in breast cancer based on WGCNA. *Onco Targets Ther.* 13, 6805–6817. doi: 10.2147/ott.s258439
- van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A. J., et al. (2018). Recovering gene interactions from single-cell data using data diffusion. *Cell* 174, 716–729.e27.
- van’t Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530–536.
- Walsh, L. A., Alvarez, M. J., Sabio, E. Y., Reingold, M., Makarov, V., Mukherjee, S., et al. (2017). An integrated systems biology approach identifies TRIM25 as a key determinant of breast cancer metastasis. *Cell Rep.* 20, 1623–1640. doi: 10.1016/j.celrep.2017.07.052
- Wang, C. C. N., Li, C. Y., Cai, J. H., Sheu, P. C., Tsai, J. J. P., Wu, M. Y., et al. (2019). Identification of prognostic candidate genes in breast cancer by integrated bioinformatic analysis. *J. Clin. Med.* 8:1160. doi: 10.3390/jcm8081160
- Wang, H. Q., Wong, H. S., Zhu, H., and Yip, T. T. (2009). A neural network-based biomarker association information extraction approach for cancer classification. *J. Biomed. Inform.* 42, 654–666. doi: 10.1016/j.jbi.2008.12.010
- Wang, J., Sun, Y., Zheng, S., Zhang, X. S., Zhou, H., and Chen, L. (2013). APG: an active protein-gene network model to quantify regulatory signals in complex biological systems. *Sci. Rep.* 3:1097.
- Wang, X. G., Peng, Y., Song, X. L., and Lan, J. P. (2016). Identification potential biomarkers and therapeutic agents in multiple myeloma based on bioinformatics analysis. *Eur. Rev. Med. Pharmacol. Sci.* 20, 810–817.
- Wang, Y., Chen, L., Ju, L., Qian, K., Liu, X., Wang, X., et al. (2019). Novel biomarkers associated with progression and prognosis of bladder cancer identified by co-expression analysis. *Front. Oncol.* 9:1030. doi: 10.3389/fonc.2019.01030
- Wang, Y., Klijn, J. G., Zhang, Y., Sieuwerts, A. M., Look, M. P., Yang, F., et al. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365, 671–679. doi: 10.1016/s0140-6736(05)17947-1
- Werhli, A. V., and Husmeier, D. (2007). Reconstructing gene regulatory networks with bayesian networks by combining expression data with multiple sources of prior knowledge. *Stat. Appl. Genet. Mol. Biol.* 6:15.
- Wilson, D. N., and Doudna, C. H. (2012). The structure and function of the eukaryotic ribosome. *Cold Spring Harb. Perspect. Biol.* 4:a011536. doi: 10.1101/cshperspect.a011536
- Yan, Z., Li, J., Xiong, Y., Xu, W., and Zheng, G. (2012). Identification of candidate colon cancer biomarkers by applying a random forest approach on microarray data. *Oncol. Rep.* 28, 1036–1042. doi: 10.3892/or.2012.1891
- Yang, Q., Wang, R., Wei, B., Peng, C., Wang, L., Hu, G., et al. (2018). Candidate biomarkers and molecular mechanism investigation for glioblastoma multiforme utilizing WGCNA. *Biomed Res. Int.* 2018:4246703.
- Ye, X., Zhang, W., Futamura, Y., and Sakurai, T. (2020). Detecting interactive gene groups for single-cell RNA-seq data based on co-expression network analysis and subgraph learning. *Cells* 9:1938. doi: 10.3390/cells9091938
- Ying, C. Y., Dominguez-Sola, D., Fabi, M., Lorenz, I. C., Hussein, S., Bansal, M., et al. (2013). MEF2B mutations lead to deregulated expression of the oncogene

- BCL6 in diffuse large B cell lymphoma. *Nat. Immunol.* 14, 1084–1092. doi: 10.1038/ni.2688
- Yu, C. Y., Xiang, S., Huang, Z., Johnson, T. S., Zhan, X., Han, Z., et al. (2019). Gene co-expression network and copy number variation analyses identify transcription factors associated with multiple myeloma progression. *Front. Genet.* 10:468. doi: 10.3389/fgene.2019.0468
- Yu, M. K., Kramer, M., Dutkowski, J., Srivas, R., Licon, K., Kreisberg, J., et al. (2016). Translation of genotype to phenotype by a hierarchy of cell subsystems. *Cell Syst.* 2, 77–88. doi: 10.1016/j.cels.2016.02.003
- Yu, M. K., Ma, J., Fisher, J., Kreisberg, J. F., Raphael, B. J., and Ideker, T. (2018). Visible machine learning for biomedicine. *Cell* 173, 1562–1565. doi: 10.1016/j.cell.2018.05.056
- Yue, W., Wang, J.-P., Conaway, M., Masamura, S., Li, Y., and Santen, R. J. (2002). Activation of the MAPK pathway enhances sensitivity of MCF-7 breast cancer cells to the mitogenic effect of estradiol. *Endocrinology* 143, 3221–3229. doi: 10.1210/en.2002-220186
- Zhan, F., Hardin, J., Kordsmeier, B., Bumm, K., Zheng, M., Tian, E., et al. (2002). Global gene expression profiling of multiple myeloma, monoclonal gammopathy of undetermined significance, and normal bone marrow plasma cells. *Blood* 99, 1745–1757. doi: 10.1182/blood.v99.5.1745
- Zhan, F., Huang, Y., Colla, S., Stewart, J. P., Hanamura, I., Gupta, S., et al. (2006). The molecular classification of multiple myeloma. *Blood* 108, 2020–2028.
- Zhan, T., Rindtorff, N., and Boutros, M. (2017). Wnt signaling in cancer. *Oncogene* 36, 1461–1473.
- Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4:17.
- Zhang, F., Chen, J., Wang, M., and Drabier, R. (2013). A neural network approach to multi-biomarker panel discovery by high-throughput plasma proteomics profiling of breast cancer. *BMC Proc.* 7(Suppl. 7):S10. doi: 10.1186/1753-6561-7-S7-S10
- Zhang, H., Yu, C. Y., Singer, B., and Xiong, M. (2001). Recursive partitioning for tumor classification with gene expression microarray data. *Proc. Natl. Acad. Sci. U.S.A.* 98, 6730–6735. doi: 10.1073/pnas.111153698
- Zhang, J., and Huang, K. (2014). Normalized lmQCM: an algorithm for detecting weak quasi-cliques in weighted graph with applications in gene co-expression module discovery in cancers. *Cancer Inform.* 13(Suppl. 3), 137–146.
- Zhang, J., Wang, L., Xu, X., Li, X., Guan, W., Meng, T., et al. (2020). Transcriptome-based network analysis unveils eight immune-related genes as molecular signatures in the immunomodulatory subtype of triple-negative breast cancer. *Front. Oncol.* 10:1787. doi: 10.3389/fonc.2020.01787
- Zhang, S., Jing, Y., Zhang, M., Zhang, Z., Ma, P., Peng, H., et al. (2015). Stroma-associated master regulators of molecular subtypes predict patient prognosis in ovarian cancer. *Sci. Rep.* 5:16066.
- Zhang, X., Yang, L., Szeto, P., Abali, G. K., Zhang, Y., Kulkarni, A., et al. (2020). The hippo pathway oncoprotein YAP promotes melanoma cell invasion and spontaneous metastasis. *Oncogene* 39, 5267–5281. doi: 10.1038/s41388-020-1362-9
- Zhao, L., Lee, B. Y., Brown, D. A., Molloy, M. P., Marx, G. M., Pavlakis, N., et al. (2009). Identification of candidate biomarkers of therapeutic response to docetaxel by proteomic profiling. *Cancer Res.* 69, 7696–7703. doi: 10.1158/0008-5472.can-08-4901

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Yu and Mitrofanova. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.