REVIEW

Applications of integrative OMICs approaches to gene regulation studies

Jing Qin[1], Bin Yan[2,3], Yaohua Hu[4], Panwen Wang[5] and Junwen Wang[5,6,*]

[1] School of Life Sciences, The Chinese University of Hong Kong, Hong Kong SAR 999077, China
[2] Laboratory for Food Safety and Environmental Technology, Institutes of Biomedicine and Biotechnology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[3] School of Biomedical Sciences, The University of Hong Kong, Hong Kong SAR 999077, China
[4] College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, China
[5] Department of Health Sciences Research and Center for Individualized Medicine, Mayo Clinic, Scottsdale, AZ 85259, USA
[6] Department of Biomedical Informatics, Arizona State University, Scottsdale, AZ 85259, USA
* Correspondence: Wang.Junwen@mayo.edu

Background: Functional genomics employs dozens of OMICs technologies to explore the functions of DNA, RNA and protein regulators in gene regulation processes. Despite each of these technologies being powerful tools on their own, like the parable of blind men and an elephant, any one single technology has a limited ability to depict the complex regulatory system. Integrative OMICS approaches have emerged and become an important area in biology and medicine. It provides a precise and effective way to study gene regulations.
Results: This article reviews current popular OMICs technologies, OMICs data integration strategies, and bioinformatics tools used for multi-dimensional data integration. We highlight the advantages of these methods, particularly in elucidating molecular basis of biological regulatory mechanisms.
Conclusions: To better understand the complexity of biological processes, we need powerful bioinformatics tools to integrate these OMICs data. Integrating multi-dimensional OMICs data will generate novel insights into system-level gene regulations and serves as a foundation for further hypothesis-driven research.

Keywords: gene regulatory networks; integrative analysis; OMICs; ChIP-seq; RNA-seq

INTRODUCTION

Precise control of gene expression is critical for development of organs and progression of diseases. Human cells employ a multi-level regulatory system to ensure their genes express at the right place and right time. In this system, various factors regulate gene expression at transcriptional, post-transcriptional, translational, post-translational and epigenetic layers. Alteration of a single or a small amount of regulatory components could disrupt the gene expression profiles, which may lead to the changes of cell phenotypes, convert cell fates or even result in diseases. Identification of the altered regulators and their downstream effects are important to uncover the molecular mechanisms in developmental processes, and to discover potential targets for treatments of diseases. To achieve this goal, OMICs technologies have been developed to measure the changes of diverse large and small molecules in a system-wide manner, including genome, transcriptome, epigenome, proteome, metabolome, interactome, etc. Despite each of these technologies being powerful tools on their own, like the parable of blind men and an elephant, any one single technology has a limited ability to depict the complex regulatory system. Thus, current strategies aim to integrate multiple types of OMICs data to investigate biological and medical issues. However, the gap between the data generation and in-depth analysis is still large. It is important to develop bioinformatics methodologies for combining multi-dimensional OMICs data, modeling diverse regulatory components systematically, and translating digital signals into biological and medical meaning.

VARIOUS OMICs TECHNOLOGIES DISSECT FUNCTIONS OF REGULATORS

The gene regulatory system, which controls genetic information flow from DNA to RNA then to proteins, performs at five major levels: signaling pathways/networks passing extracellular signals into nucleus, transcription apparatus activating or suppressing gene transcriptions, splicing factors controlling the formation of RNA isoforms, microRNAs (miRNAs) regulating mRNA and protein abundance post-transcriptionally, and long non-coding RNAs (lncRNAs) and circular RNAs (circRNAs) functioning diversely in different gene regulatory steps. Functional genomics employ dozens of OMICs technologies to explore the functions of DNA, RNA and protein regulators in the gene regulation processes. This section introduces the popular OMICs technologies applied to dissect the functions of these regulatory components in each level.

Signaling pathway/network

As shown in Figure 1, when a cell senses an extracellular signal, it passes the signal into nucleus and trigger transcription of genes in a dose-dependent manner via a chain of reactions called signaling pathway [1]. Signaling pathways involves various reactions including protein-protein interaction (PPI), phosphorylation, ubiquitination, ligand-receptor interaction, metal binding, and reactions of other small molecules. These reactions in canonical signaling pathways have been well described in databases, such as Kyoto Encyclopedia of Genes and Genomes (KEGG) and BioCarta.

However, when researchers study deeper and wider with OMICs technologies in recent years, more and more new pathways are discovered [13]. These methods mainly target two types of signaling components: proteins and small molecules (Table 1). For example, functional protein array screens direct phosphorylation activities of kinases on substrate proteins [2]. Tandem affinity purification (TAP) and immunoprecipitation (IP) pull down protein partners that interact with proteins of interests *in vivo*. When coupled with mass spectrometry (MS), TAP-MS and IP-MS identify the sequences of interacted protein partners [3,14]. Besides, MS is also applied to profile small molecule abundance [16]. Through small molecule profiling, new signaling pathways incorporating small molecules were discovered and stored in Small Molecule Pathway Database (SMPDB) [43].

These studies systematically analyze signaling transduction events mediated by proteins/phosphoproteins and small molecules, and indicate that signaling pathways are highly interconnected as signaling networks [2,44].

Discovery of new components and their interactions from OMICs data aids in extending signaling pathways and networks.

Transcription

Signaling networks may activate/deactivate transcription factors (TFs) and chromatin modifiers, which control gene transcription. TFs bound on *cis*-regulatory elements activate or repress transcription process. Chromatin modifiers control the accessibility of the DNA for TF binding, including those regulating epigenetic marks of histone modifications, DNA methylations and nucleosome positions [5]. Each of them can be measured by several omics technologies (Table 1). For instance, chromatin immunoprecipitation coupled with sequencing (ChIP-seq) or microarray (ChIP-chip) reveals the repertoire of *in vivo* protein (TF, chromatin modifier or histone) positions on the genome [17]. Bisulfite sequencing (BS-seq) is one of the most popular technologies to map DNA methylation patterns at single-base resolution [45]. Sequencing signals from these technologies mark the functional regions in a large part of non-coding genomic sequences, which were previously regarded as junk DNA. These technologies not only help to explore functional DNA elements, but also provide new insights into functional organizations of the genome and mechanisms of gene transcription [46,47].

In addition to TFs and epigenetic factors, chromatin three-dimensional structures are also critical for transcriptional activities. Distal regulatory elements, such as enhancer and silencer, can bend the DNA and connect to their targets by long-range DNA interactions. Technologies detecting long-range DNA interaction include chromatin interaction analysis by paired-end tag sequencing (ChIA-PET), circularized chromosome conformation capture (4C), carbon copy chromosome conformation capture (5C) or Hi-C [6] (Table 1). These methods not only identify the targets of distal regulatory elements, but also discover co-transcribed gene clusters that are physically connected. Together with TFs and epigenetic factors, all these DNA regulatory elements form the transcription apparatus precisely controlling gene transcriptions.

RNA editing and splicing

The products of transcription are RNA transcripts. Produced precursor RNA (pre-RNA) of protein coding genes are then edited and spliced into mature mRNAs. RNA editing is one of the mechanisms to increase the diversity of mRNAs. Enzymes change the pre-RNA sequences by insertion, deletion or deamination. RNA editing sites can be called from comparison between
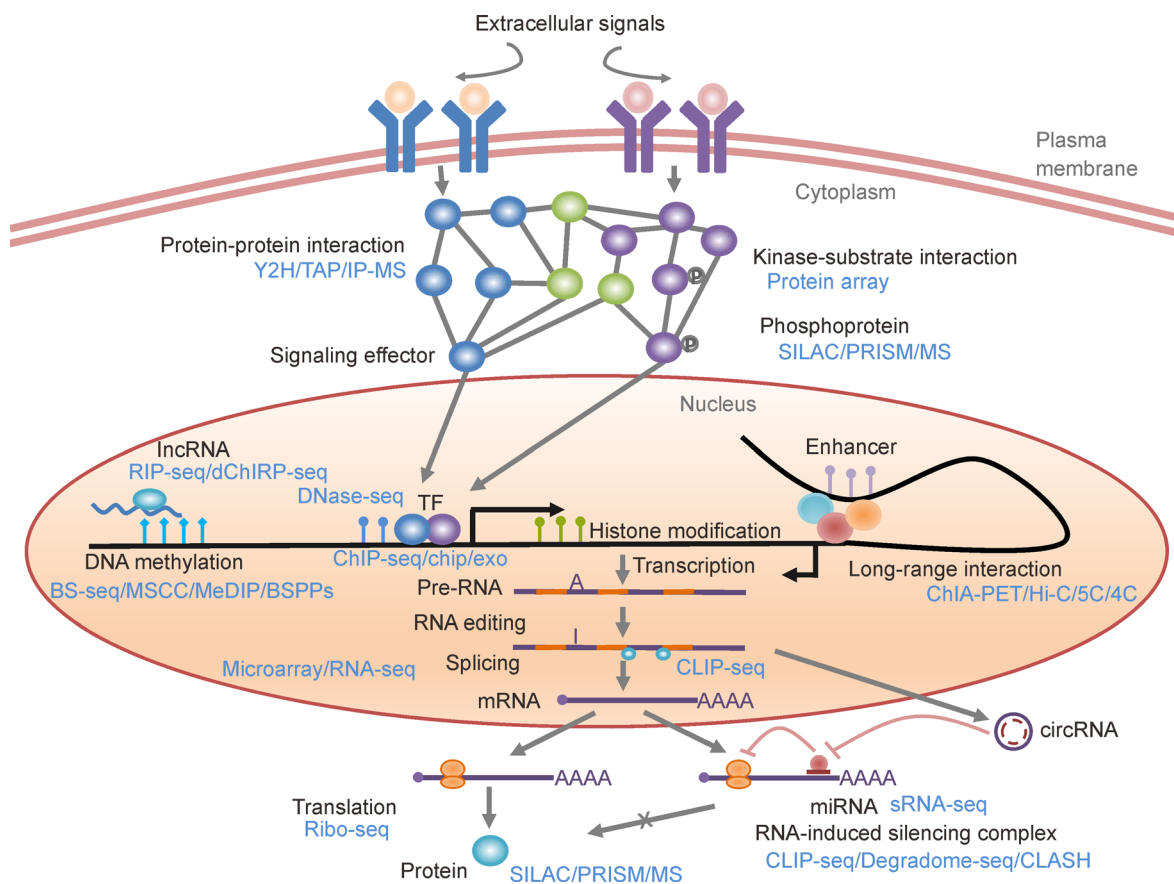
Figure 1. High-throughput OMICs data describing the complex gene regulatory system. Five parts of the gene regulatory system: **(i)** When a cell senses an extracellular signal through receptors, it passes the signal into cytoplasm via a chain of reactions, which is called signaling pathways/networks. Activated signaling effectors get into nucleus and change gene expression by binding at different subsets of targets in a dose-dependent response to the signal [1]. To identify the signaling pathways/networks, regulatory connections including protein-protein interactions (PPIs) and kinase-substrate interactions are commonly detected by several high-throughput technologies, such as yeast two-hybrid screening (Y2H), TAP, functional protein array and MS-based methods [2–4]. **(ii)** Downstream of signaling pathways/networks are the transcription networks regulated by TFs and chromatin modifiers. TFs bound on *cis*-regulatory elements can activate or repress gene transcriptions, while chromatin modifiers, including those regulate epigenetic marks of histone modifications, DNA methylations and nucleosome positions, control the accessibility of the DNA for TF binding [5]. Detecting genomic positions of TFs, chromatin modifiers and epigenetic marks is essential to infer their functions on downstream target genes. OMICs technologies for this purpose include ChIP-seq, DNase-seq, BS-seq, etc. In addition to the transcriptional and epigenetic factors, chromatin three-dimensional structures are also critical for transcription activities. Distal regulatory elements, such as enhancer and silencer, can bend the DNA and connect to their targets by long-range interactions, which can be detected by ChIA-PET, 4C, 5C or Hi-C [6]. Transcription products include mRNAs, lncRNAs, miRNAs, circRNAs, etc., whose abundance can be measured by various RNA-seq technologies. **(iii)** After transcription, pre-RNAs are edited and spliced in different ways to produce diverse mature mRNA isoforms. The profiles of different mRNA isoforms can be derived from splice-junction microarrays or RNA-seq. To understand the genome-wide principle of splicing mechanisms, CLIP-seq have been applied to map the binding sites of splicing regulators on pre-RNAs [7]. **(iv)** Mature mRNAs are subsequently exported from nucleus to cytoplasm and further translated into protein by ribosomes. As a regulator in this step, miRNAs form RISCs to regulate translation or mRNA degradation. CLIP-seq, Degradome-seq and CLASH are used to uncover the direct targets of miRNAs, while Ribo-seq that measures the ribosome occupancy along mRNAs can evaluate their effects on translation efficiency [8]. **(v)** Other important categories of regulators, such as lncRNAs and circRNAs have been found to play important roles during transcription, mRNA processing and post-transcriptional processes [9,10]. OMICs technologies detecting RNA-RNA interactions, protein-RNA interactions and DNA-RNA interactions have been utilized to discover their functions [10–12].

Table 1. **Popular OMICs technologies.**

Technology	Data type	Description
Protein array	Kinase-substrate interactome	Protein array tracks the interactions and activities of proteins [2]
IP-MS	Protein-protein interactome	Immunoprecipitation followed by mass spectrometry identifies interacting partners of a protein of interests [14]
Y2H	Protein-protein interactome	Yeast two-hybrid screening discovers protein-protein interactions [44,15]
TAP-MS	Protein-protein interactome	Tandem affinity purification combined with mass spectrometry identifies components of protein complexes [3]
MS	Metabolome	Mass spectrometry measures the consumption and release metabolites [16]
ChIP-seq/chip	Protein-DNA interactome	Chromatin immunoprecipitation coupled with sequencing or microarray reveals the repertoire of *in vivo* protein (TF, chromatin modifier, etc.) binding sites on the genome [17]
ChIP-exo	Protein-DNA interactome	Chromatin immunoprecipitation combines with the use of exonucleases to achieve a high resolution of protein binding sites [18]
BS-seq	DNA methylome	Bisulfite sequencing uses bisulfite treatment of DNA to determine its pattern of methylation [19]
MSCC	DNA methylome	Methyl sensitive cut counting, a whole genome methylation profiling method based on the sensitivity of CCGG sites to the restriction enzymes [20]
MeDIP	DNA methylome	Methylated DNA immunoprecipitation coupled with microarray or sequencing enriches methylated DNA sequences which then detected by microarray or sequencing [21]
BSPP	DNA methylome	Bisulfite padlock probes, a targeted method that isolates selected locations for methylation profiling [20]
RRBS	DNA methylome	Reduced representation bisulfite sequencing combines restriction enzymes and bisulfite sequencing to enrich high CpG sequence and measure their methylation levels [22]
Methyl-MAPS	DNA methylome	Methylation mapping analysis by paired-end sequencing [23]
DNase-seq	Open chromatin region	DNase I hypersensitive sites sequencing identifies the location of regulatory regions, based on the genome-wide sequencing of regions super sensitive to cleavage by DNase I [24]
Sono-seq	Open chromatin region	Method for isolating protein-free genomic regions by sonication of chromatin combined with a size-selection step and massively parallel short-read sequencing [25]
ATAC-seq	Open chromatin region	Assay for detecting transposase-accessible chromatin using sequencing [26]
FAIRE-Seq	Open chromatin region	Formaldehyde-assisted isolation of regulatory elements determines the sequences of regulatory region [27]
NOMe-seq	Nucleosome occupancy and methylome	A high-resolution single-molecule mapping approach to simultaneously investigate endogenous DNA methylation and nucleosome occupancies [28]
MNase-seq	Nucleosome occupancy	Paired-end sequencing of micrococcal nuclease-digested chromatin determines genome-wide nucleosome occupancy [29]
Hi-C/4C/5C	Long-range DNA interactome	Techniques to detect chromatin interactions based on chromosome conformation capture [6]
ChIA-PET	Long-range DNA interactome	Chromatin interaction analysis by paired-end tag sequencing determines long-range interactions [6]
RNA-seq	Transcriptome	RNA sequencing measures the abundance of RNA transcripts [30]
Microarray	Transcriptome	Microarray contains probes designed for RNA transcripts, on which hybridization signals indicate the abundance of RNA transcripts [31]
GRO-seq	Transcriptome	Global run-on sequencing identifies the genes that are being transcribed [32]
CLIP-seq	Protein-RNA interactome	Sequencing of RNA isolated by crosslinking immunoprecipitation maps protein-RNA binding sites *in vivo* [33]
Degradome-seq	Protein-RNA interactome	Degradome sequencing, also known as parallel analysis of RNA ends (PARE) sequencing, detects cDNA ends, which implies mRNAs that are degraded under miRNA regulations [34]

Technology	Data type	Description
RIP-seq	Protein-RNA interactome	RNA-immunoprecipitation sequencing captures the protein-bound RNAs [12]
CLASH	miRNA-RNA interactome	Crosslinking, ligation, and sequencing of hybrids directly map miRNA-RNA interactions [35]
dChIRP-seq	RNA-RNA, RNA-DNA and Protein-RNA interactome	Domain-specific chromatin isolation by RNA purification followed by sequencing dissects pairwise RNA-RNA, RNA-protein and RNA-chromatin interactions that reveal lncRNA architecture and function [11]
Structure-seq	RNA structure	A high-throughput and quantitative method detects genome-wide information on RNA structure at single-nucleotide resolution [36]
CGH	Genome	Comparative genomic hybridization surveys copy number variations [37]
DNA-seq	Genome	DNA sequencing for genome assembly or resequencing for detection of variations [38]
Exome-seq	Genome	DNA sequencing of exomes for detection of variations [39]
SILAC	Proteome	Stable isotope labeling with amino acids in cell culture quantitatively detects differences in protein abundance among samples using non-radioactive isotopic labeling [40]
PRISM	Proteome	Proteomic investigation strategy for mammals [41]
DEEP SEQ MS	Proteome	Deep efficient peptide sequencing and quantification mass spectrometry [42]
Ribo-seq	Proteome	Ribosome profiling determines the mRNAs that are being actively translated and measures the translation efficiency [8]

paired genomic DNA sequences and RNA transcripts of one sample or from RNA transcripts of multiple samples [48,49].

Then pre-RNAs are spliced in different ways and produce diverse mature mRNA isoforms. RNA splicing profiles can be derived from splice-junction microarray or RNA sequencing (RNA-seq). Splicing processes are regulated by splicing factors. They bind to pre-RNAs and regulate the selection of exons [50]. This process can be controlled by signaling pathways through post-transcriptional modification of specific splicing factors [50]. To investigate the splicing mechanism, sequencing of RNA isolated by crosslinking immunoprecipitation (CLIP-seq) uncovers the genome-wide splicing principle by mapping the binding sites of splicing factors on pre-RNAs [7].

Non-coding RNAs

A variety of RNA library preparation methods have been developed to discover functions of new RNA transcripts, such as miRNAs, lncRNAs and circRNAs, classified according to their lengths, polyadenylated/non-polyadenylated status or shapes [51,52]. These non-coding RNAs (ncRNA) are also important components in the gene regulatory system. miRNAs and argonaute (Ago) proteins form RNA-induced silencing complexes (RISCs) to regulate translation or mRNA degradation through miRNA-mRNA interactions. To identify the targets of miRNAs, CLIP-seq [33] and Degradome sequencing

(Degradome-seq, also known as PARE-seq, sequencing of parallel analysis of RNA ends [34]) were developed to detect Ago-bound mRNA fragments and miRNA directed cleavage sites at targeted mRNAs, respectively (Table 1). With the targeted mRNA sequences identified by CLIP-seq or Degradome-seq, miRNA-mRNA interactions can be predicted by computationally scanning miRNA binding sites [53]. A more powerful method, CLASH (crosslinking, ligation, and sequencing of hybrids), directly maps miRNA-mRNA interactions through an extra step that ligates miRNA and mRNA within the same RISC [35].

Another important category of non-coding RNAs, lncRNAs, have attracted great attention recently. The functions of lncRNAs were well described in other reviews [9,54]. They could be modulators of transcription, RNA processing, protein function and post-transcription process when interacting with DNA, mRNA, protein and miRNA, respectively. Technologies are applied to investigate lncRNA functions by identifying protein-RNA interactions using RNA Immunoprecipitation (RIP), or detecting RNA-RNA, RNA-DNA and protein-RNA interactions simultaneously using domain-specific chromatin isolation by RNA purification (dChIRP) coupled with sequencing (RIP-seq and dChIRP-seq) [11,12]. These technologies can discover not only the functional lncRNAs but also their targets.

circRNA is also hot research area in the recent years. Since circRNAs are circular single-stranded transcripts without poly(A) tail, circRNAs are usually enriched by

eliminating linear RNAs with RNase R and non-polyadenylated selection before RNA-seq [55]. For RNA-seq data generated without an enrichment step, circRNAs can be identified computationally by searching junction reads from back spliced exons and intron lariats [56,57]. Although the functions of most circRNAs are still unknown, a handful of circRNAs have been found to function as miRNA sponges, enhancing transcriptions or regulating RNA splicing [10]. Thus, methods detecting RNA-RNA interactions, protein-RNA interactions and DNA-RNA interactions may be also helpful in investigating circRNA functions.

CHALLENGES TO DEPICT COMPLEX GENE REGULATORY SYSTEM WITH OMICs TECHNOLOGIES

Although each OMICs technology investigates a certain type of regulators in the gene regulatory system as illustrated in the above section and Figure 1, one single method is difficult to systematically dissect the regulatory complexities. First, OMICs technologies usually experience high false positives due to their high-throughput capability. Even though many bioinformatics tools have been introduced to analyze each kind of OMICs data and apply statistical methods to rank the signals [58], they are often unable to distinguish false positives from those true ones effectively due to the inherent problems of the technology as described previously [24,59–62].

Second, results from OMICs data targeting one single type of regulators are usually obscure without considering effects of other factors. Because these regulators are not independent but work in concert to maintain biological functions, it is hard to discover complex interplays among the regulators in different layers using a single type of OMICs data. For example, the balance between transcriptional activation by a given TF and post-transcriptional suppression by miRNA forms feed forward loop or feedback loop to flexibly and precisely control target mRNA abundance [63]. Thus, mRNA abundance is dependent to not only the expression of its upstream TFs but also the concentration of miRNAs in the cytoplasm. Researchers may be unable to understand mRNA abundance changes when only either TF or miRNA data is available.

Third, multiple functions of regulators also complicate the associations between regulators and their targets observed from OMICs data. For instance, even though mRNA abundance is expected to be dependent on the expression of its upstream TFs, with some TFs that target both DNA and RNA (DNA- and RNA-binding proteins, DRBPs). Certain lncRNAs mimicking genomic DNA can compete with genomic promoter sequences and reduce the regulatory effect of TFs on transcription [64].

Similarly, miRNAs bind not only target mRNAs but also competing endogenous RNAs (ceRNAs) or circRNAs, which may compete for miRNA binding and modulate the regulatory effect of miRNAs on their genuine targets [65,66]. In these cases, unexpected weak association will be shown between the expression of regulators and their targets. Hence, the omission of these extraneous types of regulators may lead to misunderstanding of the results derived from a single OMICs analysis.

To tackle these problems, current strategies are to generate multiple types of OMICs data in a single study. Multiple types of OMICs data in a single study could cross-validate each other and reduce their false signals. Also, cross-talks between different regulatory layers could be investigated through integration of multi-level OMICs data. Thus, in the past decade, accumulated studies are incorporating multi-dimensional OMICs data. In particular, in a genome-wide scale effort to annotate the functional elements in human and mouse genomes, the ENCyclopedia Of DNA Elements (ENCODE) project detects genome-wide signals of hundreds of TFs, epigenetic markers, mRNA, ncRNA and proteins in more than one hundred cell lines or tissues [47]. Likewise, model organism ENCODE (modENCODE) contains various OMICs data of fruit fly and worm [67–69]. More specifically, the NIH Roadmap epigenomics project focusing on the function of epigenetic marks on gene transcription covers dozens of epigenetic markers in different human tissues and development stages [70].

On the other hand, to explore human genetic variations and their influence on phenotypes, the 1000 Genomes Project provides a comprehensive catalog of human genetic variations detected by sequencing of a thousand individual genomes, while RNA-seq and small RNA-seq (sRNA-seq) are also performed on the same set of individuals (http://www.1000genomes.org/). In the field of cancer biology, The Cancer Genome Atlas (TCGA) projects catalogue data from genome, transcriptome, epigenome, proteome, etc. of cancer patients [71,72]. Besides, several publicly available databases collect diverse OMICs data from different sources for a certain species or tissue/cell type [73–78]. All these resources provide an excellent opportunity to uncover the mystery of complex gene regulatory system. In the following section, we will discuss the current strategies and corresponding bioinformatics methods of OMICs data integration.

OMICs INTEGRATION STRATEGIES AND BIOINFORMATICS TOOLS

Along with the cost reduction of the OMICs technologies, rapid growth rates occur on not only the generation of

each kind of data but also the number of studies involving data integration. As shown in Figure 2, the number of studies depositing genome, transcriptome, methylation, ChIP-seq/chip, ncRNA data in Gene Expression Omnibus (GEO) and proteome data in proteomeXchange has been continuously increased since 2005. Integrative approaches have emerged almost at the same time, and have become a trend in biology and medicine. It provides a precise and effective way to elucidate the regulatory mechanisms because integrative methods highlight the interdependence of regulatory layers represented by various OMICs data and their influence over the global networks. Corresponding bioinformatics tools were also developed to meet the demands for various integrative analyses as listed in Table 2. This section will introduce current popular integration strategies and bioinformatics tools for assembling OMICs data, as well as for inferring hierarchy of gene regulatory system.

Signaling pathway/network analysis

As the most important part of signaling network, PPI and phosphorylation networks are usually detected by OMICs technologies *in vitro*. However, the detected networks usually have great amount of noise [62]. Combining transcriptome data with PPI or phosphorylation networks is the most commonly used method to reduce false positives [80]. Further, it is also an efficient approach to detect truly active signaling network by taking transcriptional regulation of downstream targets into consideration. Tools like bioPIXIE, SPINE, MINDy and ReponseNet (Table 2) use this strategy to predict active signaling and downstream transcriptional networks

simultaneously from known PPIs, genetic interactions and transcription interactions, as well as transcriptome data [81–85].

Alternatively, to detect signaling events *in vivo*, proteomics technologies are increasingly used to uncover the system-wide signaling networks. For example, large-scale perturbations of signal molecules were coupled with proteomes or phosphoproteomes [110,111]. However, these studies could not provide evidence to support the direct connections between the kinases and substrates. Integrative approach overcomes this problem by combining functional protein microarrays, phosphoproteome with bioinformatics analysis. It is able to detect the direct connections of phosphorylation network with kinase-substrate reaction activities and subsequent substrate phosphorylation status [2].

In addition to proteins, small molecules also play important roles in signaling transduction. Metabolomics targeting small molecules have also illustrated their great potential in dissecting signaling network when combined with transcriptomics [112]. However, there are only a handful of studies that have considered the regulatory effect of small molecules [113,114]. Bioinformatics methods applied to regulatory networks responded to drug treatments may also be applicable to the identification of networks downstream signaling molecules [87].

TF-gene regulation

Since RNA-seq and microarray technologies became the main approaches to measure transcriptomes, co-expression of TFs and their targets has been widely used to infer the genome-wide TF-gene regulations. The underlying
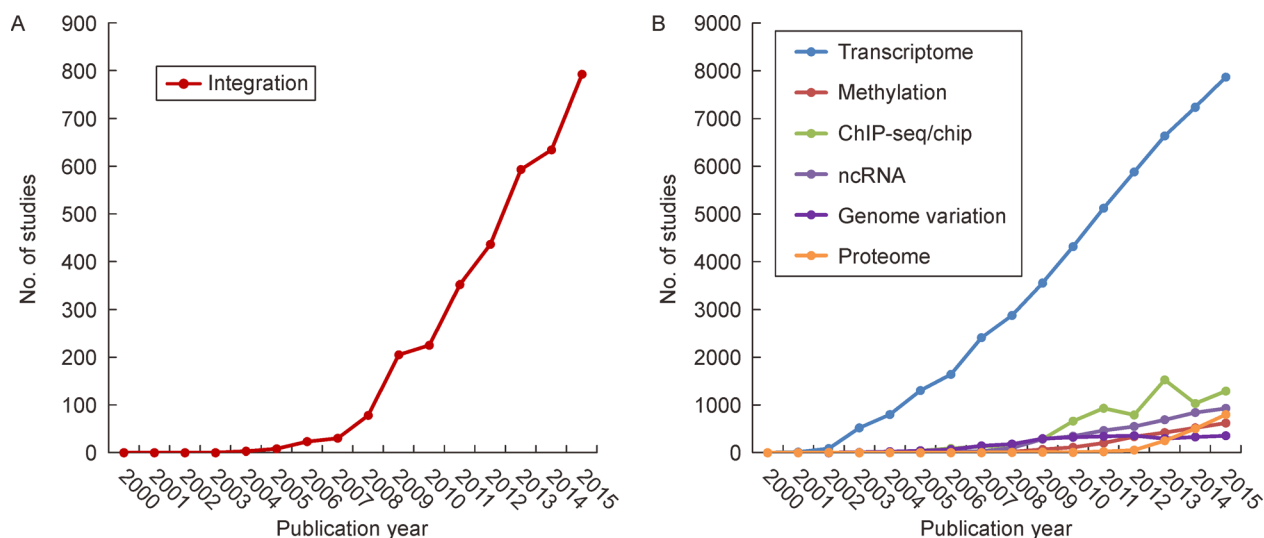

Figure 2. Statistics of OMICs studies. (A) Rapid increase of number of studies implemented with OMICs integration. (B) Numbers of studies employing different OMICs technologies.

Table 2. Current bioinformatics tools for integrative analysis of OMICs data.

Tool	Reference	Description
Signaling pathway/network analysis		
iPEAP	[79]	Integrative Pathway Enrichment Analysis Platform aggregates transcriptome, proteome, metabolome and GWAS data to detect enriched signaling pathways
PathFinder	[80]	With known PPI network, PathFinder utilizes transcriptome data, protein subcellular localization and sequence information to filter the false positives, and incorporates protein families to fix false negative pairs
bioPIXIE	[81]	An integrative system combines PPI network and gene expression data to find pathways
SPINE	[82]	Signaling-regulatory Pathway INferencE constructs PPI and protein-DNA interaction networks, and uses an integer programming solver to get final pathways from knockout expression data
ReponseNet	[83,84]	ReponseNet uses a network-optimization approach to detect both signaling and regulatory networks from the integration of genetic screen, transcriptome data and PPI networks
MINDy	[85]	Modulator inference by network dynamics (MINDy) facilitates genome-wide identification of cell-specific post-translational modulators of TF activity, which dissect the cross-talk between signaling pathways and transcriptional regulations
Zhu and Guan	[86]	A Markov chain theory is applied to detect signaling networks using a known phosphorylation networks
CEASAR	[2]	A comprehensive integrative approach incorporating functional protein microarrays, mass spectrometry and bioinformatics to detect cell-specific genome-wide phosphorylation network
SNPLS	[87]	Sparse Network-regularized Partial Least Square identifies gene-drug modules from large-scale pairwise gene-expression and drug-response data
TF-gene regulation		
ChIP-Array	[88]	A web server that integrates ChIP-seq/chip and expression data to detect direct and indirect target genes of a TF of interest
ChIP-Array 2	[89]	An enhanced version of ChIP-Array integrates long-range chromatin interaction, open chromatin region and histone modification data to dissect more comprehensive GRNs involving diverse regulatory components
BETA	[90]	Binding and Expression Target Analysis ranks the direct targets of a TF based on two probability ranking derived from ChIP-seq/chip and transcriptome data
LpRGNI	[91]	Inferring gene regulatory networks by integrating ChIP- seq/chip and transcriptome via LASSO-type regularization methods
EMBER	[92]	Expectation Maximization of Binding and Expression pRofiles, a unsupervised machine learning algorithm to search enriched gene expression pattern around TFBSs
ChIPXpress	[93]	An R package that rank TF targets by integrating ChIP-seq/chip data with large amounts of Publicly available gene Expression Data
Tang et al.	[94]	Bayesian statistical modeling and modularity analysis integrates time-series ChIP-seq and gene expression data to construct dynamic regulatory network for any given TF
Yan et al.	[95]	A bioinformatics method to uncover interactive relationships between TFs or microRNAs and genes based on matrix decomposition modeling under the joint constraints of sparseness and regulator-target connectivity
Pique-Regi et al.	[96]	A probabilistic framework that integrates epigenetic data with genomic information to draw a genome-wide map of tissue specific TFBS
miRNA-gene regulation		
GenMiR ++	[97]	GenMir ++ uses a Bayesian model to predict miRNA targets based on both mRNA 3′ UTR region sequence features and the correlation between expressions of miRNA and its targets
mirAct	[98]	A tool designed to investigate miRNA activity based on gene-expression data by using the negative regulation relationship between miRNAs and their target genes

Tool	Reference	Description
MMIA	[99]	MiRNA and MRNA Integrated Analysis is a web server that integrates miRNA and mRNA expression data with predicted miRNA target information for analyzing miRNA-associated phenotypes and biological functions
MAGIA	[100]	MiRNA and Genes Integrated Analysis is a web tool for the integrative analysis of miRNA target predictions
ProteoMirExpress	[101]	A web server that combines proteome and transcriptome data to infer miRNA-centered regulatory networks
Epigenetic regulation		
EpiRegNet	[102]	A web server that detects important histone modifications affecting the expression of genes
CMGRN	[103]	Constructing Multilevel Gene Regulatory Networks uses the Bayesian network modeling to infer causal interrelationships among transcription factors and epigenetic modifications
Multi-dimensional integration		
mirConnX	[104]	Condition-specific mRNA-miRNA network integrator uses TF binding in the promoter regions of miRNAs and mRNA, as well as predicted miRNA targets, to construct TF-miRNA-gene regulatory network
SNMNMF	[105]	A Sparse Network-Regularized Multiple Non-negative Matrix Factorization framework integrates miRNA and mRNA profiles, as well as miRNA-mRNA, TF-gene and PPI networks, for achieving modular patterns
jNMF	[106]	Joint Non-negative Matrix Factorization framework identifies multi-dimensional modules
sMBPLS	[107]	Sparse Multi-Block Partial Least Squares regression method identifies regulatory modules from multiple OMICs data
PTHGRN	[108]	Post-translational hierarchical GRN constructs multi-layer network by virtue of a graphical Gaussian model with partial least squares regression-based methodology
LRAcluster	[109]	A method using low-rank approximation based integrative probabilistic model to perform fast dimension reduction and unsupervised clustering of large-scale multi-OMICs data

assumption is that the expressions of a TF and its target are correlated. However, even though mRNA abundance represents the expression of a TF, it may not reflect its activity in the regulatory processes, which is determined by the dynamic binding of TF proteins on the regulatory elements of targets. Several factors, such as miRNAs at post-transcriptional regulation level, post-translational modification (PTM), accessibility of the DNA regulatory element and affinity of the TF-DNA interaction, may affect the functional activity of the TF. Conversely, a high correlation of two genes could be due to other co-regulatory relationship or even random association rather than TF regulation. Therefore, the accuracy of TF-gene regulation inferred from transcriptome data alone is unsatisfactory, although various statistical and mathematical methods have been introduced to reduce false positives [115].

In ChIP-seq/chip analysis, *in vivo* TF activities in binding to the targeted DNA sequence are measured. A proximal gene of each binding site is usually considered to be the target of a TF. However, ChIP-seq/chip alone does not provide the response of target genes to TF binding. The expression changes of potential target genes of a TF could be identified by comparing the transcriptome before and after a perturbation (knockdown or overexpression) of the TF. By comparing the results from the transcriptomes and ChIP-seq/chip in the same condition, it has been reported that, only 3% of targets identified by ChIP-seq/chip have expression changes after TF perturbation, and only 3% of genes with expression changes after TF perturbation are adjacent to ChIP-seq/chip peaks in yeast [116]. In mammalians, the proportion of genes is about 6%–17% [88]. This phenomenon of unaffected targets may be explained by several reasons: TFs may bind on distal enhancers, thus the proximal genes are not the targets; TF binding on targets initiates the transcription, but elongation may be regulated by other TFs; and expression of most targets without TF binding on their promoters may be indirectly affected by TF perturbation. Thus, ChIP-seq/chip data together with transcriptome data under TF perturbation are necessary to identify the true targets of a TF.

With two types of data, a direct target of a TF is usually determined as a gene that contains active TF binding sites on its promoter and is differentially expressed following the perturbation. However, given thousands of TFs in a species, it would be an arduous task to search all active targeted sites of all TFs using ChIP-seq. Technologies that detect open chromatin regions or nucleosome positions, such as DNase I hypersensitive sites sequencing (DNase-seq), may discover all active TF binding elements in a certain condition [117,118]. It provides an opportunity to assay genome-wide binding of many TFs in a single experiment [96]. Compared to ChIP-seq, paired DNase-seq with transcriptome data has great advantages in constructing more comprehensive gene regulatory network (GRN) [119].

Furthermore, both TF binding and open chromatin sites indicate not only proximal but also distal regulatory elements. The distal regulatory elements connect to their target genes through long-range DNA interactions. Combinatory analysis of genome-wide long-range chromatin interaction and ChIP-seq data elucidate the function of distal regulatory elements and the impact on gene regulations [120]. Incorporation of long-range DNA interactome improves the TF-GRN inference by adding direct evidence for the TFs that bind on distal elements of their true targets [89].

miRNA-gene regulatory modules

Under post-transcriptional level, a typical function of miRNA is to regulate the abundance of mRNAs, protein or both of its target gene products through translational repression and/or RNA degradation. In miRNA studies, genome-wide expression profiles of miRNA and mRNA are generated by sRNA-seq and RNA-seq, respectively. With this data, bioinformatics tools can infer miRNA targets by using the negative correlation between mRNA and miRNA expression (Table 2). To distinguish the direct and indirect targets of miRNAs, most of these tools incorporate predicted miRNA binding site information.

However, miRNA-mRNA binding prediction programs are usually biased. They favor canonical binding sites, which contain exact match seed region and are located in 3′ untranslated region (3′ UTR). Even though CLIP-seq and Degradome-seq can detect targeted mRNAs *in vivo*, the miRNA-mRNA interactions are also biased because the exact identification of miRNA binding sites still rely on predictions. With unbiased method, CLASH, non-canonical miRNA binding sites with mismatch or located in other regions of mRNAs were also found to associate with mRNA down-regulation [35]. Thus, bioinformatics tool considering transcriptomes coupled with miRNA binding sites information detected by CLASH may improve the sensitivity of miRNA target identification.

Furthermore, these approaches do not consider the miRNA effects on the target protein abundance, even though these proteins are the final effectors of miRNA actions. Measuring mRNA abundance may not be an ideal method for identifying miRNA targets as mRNA levels do not necessarily correlate with protein expression [121–123]. Thus, using quantitative proteomics strategies has emerged as a key technique for the identification of miRNA targets. Recent development of proteomics technology makes it reach a much higher protein sequencing coverage comparable to RNA-seq [42]. Despite this obvious advantage in use of proteome data, to date, only a few computational methods have incorporated proteome data into miRNA analysis [101].

Besides repression on translation, similar to TFs, miRNAs are also found to switch from down-regulation to up-regulation of translation in different conditions [124,125]. Thus, current bioinformatics tools considering only the negative regulatory relationships between miRNA and their targets might loss many up-regulated targets. Ribosomal profiling (Ribo-Seq) is a highly promising technique to assess the effect of miRNAs on translational regulation [8]. It represents an even more direct and accurate measurement for translation efficiency than proteomics technologies, since protein abundance is affected by turnover rate. Combining ribosomal profiles and miRNA profiles, as well as miRNA binding information from CLASH, could be a good strategy to comprehend the complexity of miRNA functions. However, corresponding bioinformatics tool has not been available yet. Moreover, more bioinformatics analyses are necessary to assess different combinations of these data types for achieving an optimal integrative strategy for miRNA studies.

Integration of epigenetic data

Combining the genome-wide profiles of multiple histone modifications, researchers are able to develop mathematical models to predict gene expression level, enhancer-templated ncRNA abundance and enhancer-promoter interactions [126–131]. Thus, epigenome adds one more regulatory layer to describe the GRN. With the histone modification profiles and transcriptome data, bioinformatics tools can build an epigenetic regulatory network, unraveling collaborations among TFs and epigenetic modifications on transcriptional regulation [102,103]. In addition to TFs, lncRNAs are also interacted with epigenetic factor and play important functions in epigenetic regulatory network. Integration of RIP-chip, epigenome and transcriptome were applied to investigate the lncRNAs-epigenetic network [132]. However, it still lacks bioinformatics tool for such integrative analysis currently.

Methods for multi-dimensional integration

Although above partial integrations of OMICs data show the advantages in improving accuracy of regulatory function identification, they usually focus on single regulatory layer. Recently, an increasing number of studies integrate multi-dimensional OMICs to investigate complex gene regulatory hierarchies. When multiple types of OMICs data are involved, the gene regulatory system is usually modeled as multi-layered GRNs. Numerous bioinformatics tools and methods have been developed to cater for the demands of multi-dimensional data integration.

The most studied networks are transcriptional-post-transcriptional regulatory networks. For example, the condition-specific mRNA-miRNA network integrator (mirConnX) uses TF binding in the promoter regions of miRNAs and mRNAs, as well as predicted miRNA targets, to construct TF-miRNA-gene regulatory networks through statistical association measure [104]. Employing a similar approach to cancer biology, Knouf et al., combined TF binding and the transcriptome data for both mRNA and miRNA to search new regulatory interactions between TFs and miRNAs that are aberrant in cancer samples [133]. Besides statistical methods, mathematical modeling is also used to solve this problem [105].

To add more regulatory layers onto the GRNs, studies incorporate genome and epigenome data [134,135]. Then, multi-dimensional data are often formulated in machine-learning frameworks for regulatory network analysis [136]. In particular, Zhang et al. in 2012 applied the joint matrix factorization technique to integrate DNA methylation, mRNA and miRNA expression data of ovarian cancer samples from TCGA project and to identify regulatory modules active in ovarian cancer [106]. Similar analysis could also be performed using a sparse Multi-Block Partial Least Squares (sMBPLS) regression method [107]. On the other hand, by considering the regulatory relationships between different types of regulators and their targets, Sintupisut et al. used pair-wise association studies to screen associated regulatory module for each molecular aberration in glioblastoma multiforme [137]. Followed by module merging, identified molecular characteristics showed strong prognostic power [137]. These studies indicate that OMICs integration can find cancer regulatory modules that would be overlooked with only a single type of data.

Furthermore, linking the signaling network to TF/epigenetic modification-gene network is able to discovery the cross-talks among PTMs, TFs and epigenetic modifications in transcription regulation. A web server, post-translational hierarchical GRN (PTHGRN) constructs this multi-layer network by virtue of a graphical Gaussian model with partial least squares regression-based methodology [108]. The development of these bioinformatics tools for the multi-dimensional OMICs analysis provides unprecedented opportunities to dissect the cross-layer regulatory interplay in various biological and medical studies.

INTEGRATING MULTIPLE OMICs TO ADVANCED BIOLOGY AND MEDICINE

Understanding the gene regulatory system from integration of OMICs data has profoundly changed the strategy on basic biological research and is playing significant roles in medical fields (Figure 3). Before OMICs technologies are developed, results from traditional experiments usually did not realize the expectation due to the interactions of a large number of unknown factors, so the initial hypothesis needed to be continually modified and validated which may cost a lot of time and efforts. In the beginning of OMICs development, limited by the budget, only single OMICs experiment was designed for an original hypothesis. Then, major efforts were contributed to data generation and analysis. Due to the high false positive, results from the OMICs data required further validation using traditional experiments. In post-OMICs era, data generation is no longer a problem because the costs of high-throughput technologies have been continually reduced. Mass of public OMICs data could be first integrated and give rise to a biological/medical hypothesis. Then more OMICs data could be generated to test the hypothesis. However, major efforts are required to combine multi-dimensional OMICs data, model diverse regulatory components systematically, and interpret digital results into biological and medical contexts [138].

Integrative OMICs approaches help comprehensive understanding of transcriptional and epigenetic dynamics in stem cell research

OMICs technologies have been extensively used in the stem cell research area. Since 2005, by combining genomic occupancy identification and transcriptome profiling, Young et al. have identified core regulatory circuits in embryonic stem cells (ESCs), which cover key TFs [139], miRNAs [140] and epigenetic regulators [141] that are all crucial for ESC maintenance. Some of these key factors have been successfully applied to reprogram mature somatic cells into ESC-like induced pluripotent stem cells (iPSCs) later [142,143].

This technology has led to the revolution of regeneration medicine. Patient-specific iPSCs with the potential to differentiate into any type of cells potentiate various clinical applications for many diseases, including drug screening and cell-replacement therapy without immune
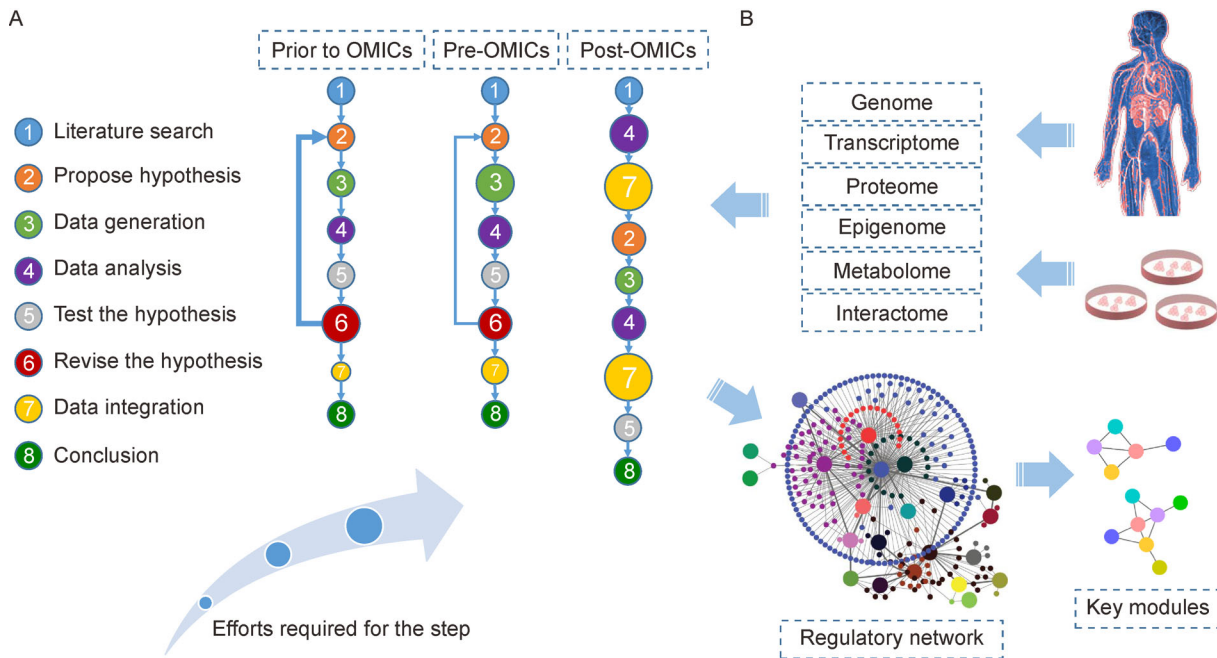
Figure 3. **Integrative OMICs have profoundly changed the strategy on basic biological research and is playing significant roles in medical fields.** (A) Before OMICs technologies are developed, results from traditional experiments usually did not realize the expectation, so the initial hypothesis needed to be continually modified and validated. In the beginning of OMICs development, only single OMICs experiment was performed, and major efforts were contributed to data generation and analysis. In post-OMICs era, data generation is no longer a problem, but major efforts are required to combine multi-dimensional OMICs data. (B) Integration of multidimensional OMICs profiles provides a more precise and effective methodology to comprehend the gene regulatory mechanisms and identify key factors/modules underlying the biological or medical processes of interests.

rejection issues [144]. However, clinical applications of the iPSCs are still restricted by the limited purity and yield, and the potential hazard of tumor development. Further basic research is required to understand the transcriptional and epigenetic key events in the processes of reprogramming through diverse integration of OMICs data [145]. This basic research helps to improve the reprogramming protocols and moving this technology closer to clinical application. Similarly, to refine the cell differentiation strategies that are also necessary for iPSC-based clinical applications, researchers have combined BS-seq, ChIP-seq and RNA-seq to investigate the transcriptional and epigenetic dynamics during the differentiations of various cell lineages [146–149]. These studies identified diverse signatures to guide the differentiation protocols.

Furthermore, in addition to iPSC-based cell engineering, direct cell fate conversion between differentiated cells (transdifferentiation) is an alternative technique with higher conversion efficiency, lower risk of tumorigenecity and a generally shorter reprogramming phase [150]. The major challenge facing transdifferentiation is how to

determine the TF combination for converting one cell type into another. Shmulevich and colleagues used expression rank difference to identify TFs that may control cell lineage specifications [151]. Though some of these TFs show potential applications for cell fate conversion, without considering their regulatory functions on lineage-specific genes, many identified cell-specific TFs may not be the determining factors for their target cell identity. Integrative OMICs approaches contributed to solving this problem by connecting TFs to lineage-specific genes through network-based methods that incorporate transcriptomes and genome-wide information of TF-DNA interactions [152] or PPIs [153]. Comprehensive computational predictions of key TFs for cell fate conversions of over hundreds of cell types could open the door for experimental biologists to create any type of cells they need through transdifferentiation.

Integrative OMICs methods identify gene regulatory disorders in complex diseases

Genetic variants may influence complex traits by altering

amino acid sequences of protein coding genes or other functional elements that modulate gene expression. Genome-wide association studies (GWASs) map genetic variations associated with various complex diseases, like cancers, diabetes and obesity [154]. Although disease-associated genetic variants within coding regions have been well studied, the functions of most genetic variants located in non-coding regions are still unclear. Integrative OMICs methods can be used to identify regulatory genetic variants causing erroneous gene regulation in complex diseases, where single OMICs analysis fails.

For example, due to the indirect linkage through a complex multi-layer gene regulatory system, the statistical significance of associations between regulatory genetic variants and complex traits or diseases is usually weaker than those of nonsynonymous genetic variants that directly affect protein functions [155]. To tackle this problem, a combination of genomics, transcriptomics and functional genomics, maps genetic variants associated with regulatory quantitative traits, such as gene expression (expression quantitative trait locus, eQTL), DNA methylation (methylation quantitative trait locus, mQTL), alternative splicing (splicing quantitative trait locus, sQTL) [156]. These methods dissect regulatory complexities between genetic variants and disease phenotypes that describe the functional linkages among genetic variants, regulators, gene expression and disease phenotypes, which in turn, help to uncover regulatory genetic variants leading to regulatory disorders and key factors causing disease phenotypes.

In addition to QTL mapping methods, researchers combine comprehensive OMICs data sets to explore the underlying effects of disease-associated variations on erroneous gene regulation. For instance, onco-proteogenomics investigates the effects of protein-coding variations on gene regulatory system with patient genomes, transcriptomes and proteomes [157]. Onco-proteogenomics built customized databases of peptides deduced from patient genomes and transcriptomes to improve the proteomic detection of cancer-specific peptides. Identified variations in genomes, transcriptomes and proteomes describe the altered information flow from DNA to proteins. Since proteins are central to cell function, incorporating proteome facilitate identification of changes in signaling pathways and PTMs, which improve understanding of how gene networks are dysregulated.

Integration of OMICs data can also improve clinical applications, such as classification of disease subtypes and prediction of disease phenotypes. It has been found in several studies that new cancer subtypes can be identified by molecular characterization of cancer patients with comprehensive OMICs data types [72,158,159]. To assess the application of OMICs data to survival prediction, Yuan *et al*. compares the prognostic power of diverse OMICs data (somatic copy-number alteration, DNA methylation and mRNA, microRNA and protein expression) alone and in combination with clinical data of four cancer types [160]. Incorporating OMICs data with clinical variables have shown more efficient predictions of patient survival in three cancers when using two different prediction models.

However, applications of integrative OMICs in complex diseases are always limited by the small effective sample sizes. It leads to high-dimensional analysis problems, where the dimension of the data vectors is much larger than the sample size. To deal with such high-dimensional analysis problems, dimension reduction is a promising strategy to improve the analysis [109]. On the other hand, increasing the effective sample size is the other way to ensure the accuracy of integrative OMICs analysis. But integrative OMICs analysis requires complete set of data matrix for all patient samples, which are usually hampered by specimen availability and cost. Fortunately, computational imputations for missing data can be applied to complete the data matrix so that it increases the effective sample size. Basically, imputation methods can use available information to approximate the unknown missing values when those valuables are correlated. For example, Ernst and Kellis imputed 25 types of OMICs signals (histone marks, DNA accessibility, DNA methylation, RNA-seq, etc.) for 127 human tissues/cell types where only 26% of signals have been profiled by OMICs experiments [161]. By comparing imputed and experimentally derived epigenomes of the same sample, Ernst and Kellis reported that imputation methods were able to recover more than 90% signal peaks for most epigenetic marks (Figure 2 in [161]). Thus, when a full set of multi-dimensional OMICs data have been measured for a small set of reference individuals, we might be able to impute the missing OMICs data for a larger set of individuals, for whom only a subset of OMICs data or reduced data points are available. This method has been applied to impute the genotypes missing in SNP arrays with a reference set of whole genome sequencing data [162]. To integrate association between genomic, transcriptomic variations and disease phenotypes, Gusev *et al*. imputed gene expression values from genome data of patients whose transcriptomes were not profiled [163]. With computational imputation, effective sample size can reach up to tens of thousands of individuals, which dramatically improve the performance of statistical tests.

Another challenge of integrative OMICs analysis is how to merge and normalized signals from diverse platforms and data sources. Most of the current normalization methods were developed for signals from the same platform [164–167]. Methods that eliminate batch effects of data from different studies or research groups were also

designed for one single platform [168,169]. In integrative analysis, signals from different platforms may have different technical bias. Normalization between platforms is critical for downstream analysis. To tackle this problem, rank based approaches and matrix normalization have been adopted for cross-platform analyses [170–172]. These methods match different identifiers of different platforms, merge datasets using different scales of measurements, and correct data bias.

CONCLUSION

With increasing attention focusing on OMICs integration approaches, even though many challenges related to in-depth analysis remain, researchers continue to work towards the ultimate goals of employing these approaches for disclosing the mystery of complex GRNs and further advancing basic studies and translational medicine. This review summarizes current experimental and bioinformatics methods and demonstrates that an integrative data analysis goes beyond the output we can achieve from a single data analysis. Currently, most of bioinformatics tools are designed for only a few data types. Many OMICs data listed in Table 1 have not been considered. In the future, with the development of new and different combinations of OMICs data becoming available, more powerful mathematical/statistical models are required to quantitatively describe each component in the system and the interactions among them based on the measurements of the OMICs technologies. Bioinformatics analysis should also assess and compare different combinations of OMICs data, which would feedback to the OMICs experiment design. Advancements of OMICs technologies and bioinformatics analyses promote and benefit each other. Integrative OMICs data interpretation will generate new insights into system-wide gene regulation and serve as a foundation for further hypothesis-driven investigations.

COMPLIANCE WITH ETHICS GUIDELINES

The authors Jing Qin, Bin Yan, Yaohua Hu, Panwen Wang and Junwen Wang declare that they have no conflict of interests.

This article does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

1. Lee, K. L., Lim, S. K., Orlov, Y. L., Yit, Y., Yang, H., Ang, L. T., Poellinger, L. and Lim, B. (2011) Graded Nodal/Activin signaling titrates conversion of quantitative phospho-Smad2 levels into qualitative embryonic stem cell fate decisions. PLoS Genet., 7, e1002130

2. Newman, R. H., Hu, J., Rho, H. S., Xie, Z., Woodard, C., Neiswinger, J., Cooper, C., Shirley, M., Clark, H. M., Hu, S., *et al.* (2013) Construction of human activity-based phosphorylation networks. Mol. Syst. Biol., 9, 655

3. Gavin, A. C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., *et al.* (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature, 415, 141–147

4. Fields, S. and Song, O. (1989) A novel genetic system to detect protein-protein interactions. Nature, 340, 245–246

5. Chen, T. and Dent, S. Y. (2014) Chromatin modifiers and remodellers: regulators of cellular differentiation. Nat. Rev. Genet., 15, 93–106

6. Dekker, J., Marti-Renom, M. A. and Mirny, L. A. (2013) Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat. Rev. Genet., 14, 390–403

7. Witten, J. T. and Ule, J. (2011) Understanding splicing regulation through RNA splicing maps. Trends Genet., 27, 89–97

8. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. and Weissman, J. S. (2009) Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. Science, 324, 218–223

9. Geisler, S. and Coller, J. (2013) RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. Nat. Rev. Mol. Cell Biol., 14, 699–712

10. Chen, L. L. (2016) The biogenesis and emerging roles of circular RNAs. Nat. Rev. Mol. Cell Biol., 17, 205–211

11. Quinn, J. J., Ilik, I. A., Qu, K., Georgiev, P., Chu, C., Akhtar, A. and Chang, H. Y. (2014) Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. Nat. Biotechnol., 32, 933–940

12. Di Ruscio, A., Ebralidze, A. K., Benoukraf, T., Amabile, G., Goff, L. A., Terragni, J., Figueroa, M. E., De Figueiredo Pontes, L. L., Alberich-Jorda, M., Zhang, P., *et al.* (2013) DNMT1-interacting RNAs block gene-specific DNA methylation. Nature, 503, 371–376

13. Gómez-Orte, E., Sáenz-Narciso, B., Moreno, S. and Cabello, J. (2013) Multiple functions of the noncanonical Wnt pathway. Trends Genet., 29, 545–553

14. Liang, J., Wan, M., Zhang, Y., Gu, P., Xin, H., Jung, S. Y., Qin, J., Wong, J., Cooney, A. J., Liu, D., *et al.* (2008) Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. Nat. Cell Biol., 10, 731–739

15. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M. and Sakaki, Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. USA, 98, 4569–4574

16. Jain, M., Nilsson, R., Sharma, S., Madhusudhan, N., Kitami, T., Souza, A. L., Kafri, R., Kirschner, M. W., Clish, C. B. and Mootha, V. K. (2012) Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. Science, 336, 1040–1044

17. Park, P. J. (2009) ChIP-seq: advantages and challenges of a maturing technology. Nat. Rev. Genet., 10, 669–680

18. Rhee, H. S. and Pugh, B. F. (2012) ChIP-exo method for identifying

genomic location of DNA-binding proteins with near-single-nucleotide accuracy. In Current Protocols In Molecular Biology, Chapter 21, Unit 21–24. Wiley

19. Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q. M., *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature, 462, 315–322

20. Ball, M. P., Li, J. B., Gao, Y., Lee, J. H., LeProust, E. M., Park, I. H., Xie, B., Daley, G. Q. and Church, G. M. (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat. Biotechnol., 27, 361–368

21. Pelizzola, M., Koga, Y., Urban, A. E., Krauthammer, M., Weissman, S., Halaban, R. and Molinaro, A. M. (2008) MEDME: an experimental and analytical methodology for the estimation of DNA methylation levels based on microarray derived MeDIP-enrichment. Genome Res., 18, 1652–1659

22. Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S. and Jaenisch, R. (2005) Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res., 33, 5868–5877

23. Edwards, J. R., O'Donnell, A. H., Rollins, R. A., Peckham, H. E., Lee, C., Milekic, M. H., Chanrion, B., Fu, Y., Su, T., Hibshoosh, H., *et al.* (2010) Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. Genome Res., 20, 972–980

24. He, H. H., Meyer, C. A., Hu, S. S., Chen, M. W., Zang, C., Liu, Y., Rao, P. K., Fei, T., Xu, H., Long, H., *et al.* (2014) Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. Nat. Methods, 11, 73–78

25. Auerbach, R. K., Euskirchen, G., Rozowsky, J., Lamarre-Vincent, N., Moqtaderi, Z., Lefrançois, P., Struhl, K., Gerstein, M. and Snyder, M. (2009) Mapping accessible chromatin regions using Sono-Seq. Proc. Natl. Acad. Sci. USA, 106, 14926–14931

26. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. and Greenleaf, W. J. (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods, 10, 1213–1218

27. Gaulton, K. J., Nammo, T., Pasquali, L., Simon, J. M., Giresi, P. G., Fogarty, M. P., Panhuis, T. M., Mieczkowski, P., Secchi, A., Bosco, D., *et al.* (2010) A map of open chromatin in human pancreatic islets. Nat. Genet., 42, 255–259

28. You, J. S., Kelly, T. K., De Carvalho, D. D., Taberlay, P. C., Liang, G. and Jones, P. A. (2011) OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes. Proc. Natl. Acad. Sci. USA, 108, 14497–14502

29. Schones, D. E., Cui, K., Cuddapah, S., Roh, T. Y., Barski, A., Wang, Z., Wei, G. and Zhao, K. (2008) Dynamic regulation of nucleosome positioning in the human genome. Cell, 132, 887–898

30. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods, 5, 621–628

31. Schena, M., Shalon, D., Davis, R. W. and Brown, P. O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science, 270, 467–470

32. Core, L. J., Waterfall, J. J. and Lis, J. T. (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science, 322, 1845–1848

33. Chi, S. W., Zang, J. B., Mele, A. and Darnell, R. B. (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature,

460, 479–486

34. German, M. A., Pillay, M., Jeong, D. H., Hetawal, A., Luo, S., Janardhanan, P., Kannan, V., Rymarquis, L. A., Nobuta, K., German, R., *et al.* (2008) Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. Nat. Biotechnol., 26, 941–946

35. Helwak, A., Kudla, G., Dudnakova, T. and Tollervey, D. (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell, 153, 654–665

36. Ding, Y., Tang, Y., Kwok, C. K., Zhang, Y., Bevilacqua, P. C. and Assmann, S. M. (2014) *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature, 505, 696–700

37. Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., *et al.* (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat. Genet., 20, 207–211

38. Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., *et al.* (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature, 456, 53–59

39. Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E. E., *et al.* (2009) Targeted capture and massively parallel sequencing of 12 human exomes. Nature, 461, 272–276

40. Krüger, M., Moser, M., Ussar, S., Thievessen, I., Luber, C. A., Forner, F., Schmidt, S., Zanivan, S., Füssler, R. and Mann, M. (2008) SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. Cell, 134, 353–364

41. Kislinger, T., Rahman, K., Radulovic, D., Cox, B., Rossant, J. and Emili, A. (2003) PRISM, a generic large scale proteomic investigation strategy for mammals. Mol. Cell. Proteomics, 2, 96–106

42. Zhou, F., Lu, Y., Ficarro, S. B., Adelmant, G., Jiang, W., Luckey, C. J. and Marto, J. A. (2013) Genome-scale proteome quantification by DEEP SEQ mass spectrometry. Nat. Commun., 4, 2171

43. Jewison, T., Su, Y., Disfany, F. M., Liang, Y., Knox, C., Maciejewski, A., Poelzer, J., Huynh, J., Zhou, Y., Arndt, D., *et al.*(2014) SMPDB 2.0: big improvements to the Small Molecule Pathway Database. Nucleic Acids Res., 42, D478–D484

44. Song, C., Ye, M., Liu, Z., Cheng, H., Jiang, X., Han, G., Songyang, Z., Tan, Y., Wang, H., Ren, J., *et al.* (2012) Systematic analysis of protein phosphorylation networks from phosphoproteomic data. Mol. Cell. Proteomics, 11, 1070–1083

45. Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., *et al.* (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature, 454, 766–770

46. Neph, S., Vierstra, J., Stergachis, A. B., Reynolds, A. P., Haugen, E., Vernot, B., Thurman, R. E., John, S., Sandstrom, R., Johnson, A. K., *et al.* (2012) An expansive human regulatory lexicon encoded in transcription factor footprints. Nature, 489, 83–90

47. Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., *et al.* (2012) An integrated encyclopedia of DNA elements in the human genome. Nature, 489, 57–74

48. Ramaswami, G., Zhang, R., Piskol, R., Keegan, L. P., Deng, P., O'Connell, M. A. and Li, J. B. (2013) Identifying RNA editing sites using RNA sequencing data alone. Nat. Methods, 10, 128–132

49. Ramaswami, G., Lin, W., Piskol, R., Tan, M. H., Davis, C. and Li, J.

B. (2012) Accurate identification of human Alu and non-Alu RNA editing sites. Nat. Methods, 9, 579–581

50. Fu, X. D. and Ares, M. Jr. (2014) Context-dependent control of alternative splicing by RNA-binding proteins. Nat. Rev. Genet., 15, 689–701

51. Wang, Z., Gerstein, M. and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet., 10, 57–63

52. Yang, L., Duff, M. O., Graveley, B. R., Carmichael, G. G. and Chen, L. L. (2011) Genomewide characterization of non-polyadenylated RNAs. Genome Biol., 12, R16

53. Yang, J. H., Li, J. H., Shao, P., Zhou, H., Chen, Y. Q. and Qu, L. H. (2011) starBase: a database for exploring microRNA–mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. Nucleic Acids Res., 39, D202–D209

54. Lau, E. (2014) Non-coding RNA: zooming in on lncRNA functions. Nat. Rev. Genet., 15, 574–575

55. Venø, M. T., Hansen, T. B., Venø, S. T., Clausen, B. H., Grebing, M., Finsen, B., Holm, I. E. and Kjems, J. (2015) Spatio-temporal regulation of circular RNA expression during porcine embryonic brain development. Genome Biol., 16, 245

56. Gao, Y., Wang, J. and Zhao, F. (2015) CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. Genome Biol., 16, 4

57. Salzman, J., Gawad, C., Wang, P. L., Lacayo, N. and Brown, P. O. (2012) Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. PLoS One, 7, e30733

58. Qin, Y., Yalamanchili, H.K., Qin, J., Yan, B. and Wang, J. (2015) The current status and challenges in computational analysis of genomic big data. Big data research, 2, 12–18

59. Kluger, Y., Yu, H., Qian, J. and Gerstein, M. (2003) Relationship between gene co-expression and probe localization on microarray slides. BMC Genomics, 4, 49

60. Robasky, K., Lewis, N. E. and Church, G. M. (2014) The role of replicates for error mitigation in next-generation sequencing. Nat. Rev. Genet., 15, 56–62

61. Teytelman, L., Thurtle, D. M., Rine, J. and van Oudenaarden, A. (2013) Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. Proc. Natl. Acad. Sci. USA, 110, 18602–18607

62. von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S. and Bork, P. (2002) Comparative assessment of large-scale data sets of protein-protein interactions. Nature, 417, 399–403

63. El Gazzar, M. and McCall, C. E. (2012) MicroRNAs regulatory networks in myeloid lineage development and differentiation: regulators of the regulators. Immunol. Cell Biol., 90, 587–593

64. Hudson, W. H. and Ortlund, E. A. (2014) The structure, function and evolution of proteins that bind DNA and RNA. Nat. Rev. Mol. Cell Biol., 15, 749–760

65. Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W. J. and Pandolfi, P. P. (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature, 465, 1033–1038

66. Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K. and Kjems, J. (2013) Natural RNA circles function as efficient microRNA sponges. Nature, 495, 384–388

67. modEncode Consortium, Roy, S., Ernst, J.J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C. A., Ma, L. et al. (2010) Identification of functional elements and

regulatory circuits by Drosophila modENCODE. Science, 330, 1787–1797

68. Nègre, N., Brown, C. D., Ma, L., Bristow, C. A., Miller, S. W., Wagner, U., Kheradpour, P., Eaton, M. L., Loriaux, P., Sealfon, R., et al. (2011) A cis-regulatory map of the Drosophila genome. Nature, 471, 527–531

69. Gerstein, M. B., Lu, Z. J., Van Nostrand, E. L., Cheng, C., Arshinoff, B. I., Liu, T., Yip, K. Y., Robilotto, R., Rechtsteiner, A., Ikegami, K., et al. (2010) Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science, 330, 1775–1787

70. Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J., et al. (2015) Integrative analysis of 111 reference human epigenomes. Nature, 518, 317–330

71. Bell, D., Berchuck, A., Birrer, M., Chien, J., Cramer, D. W., Dao, F., Dhir, R., DiSaia, P., Gabra, H., Glenn, P., et al. (2011) Integrated genomic analyses of ovarian carcinoma. Nature, 474, 609–615

72. Bass, A. J., Thorsson, V., Shmulevich, I., Reynolds, S. M., Miller, M., Bernard, B., Hinoue, T., Laird, P. W., Curtis, C., Shen, H., et al. (2014) Comprehensive molecular characterization of gastric adenocarcinoma. Nature, 513, 202–209

73. Xu, H., Baroukh, C., Dannenfelser, R., Chen, E.Y., Tan, C. M., Kou, Y., Kim, Y. E., Lemischka, I. R. and Ma'ayan, A. (2013) ESCAPE: database for integrating high-content published data collected from human and mouse embryonic stem cells. Database, 2013, bat045

74. Marygold, S. J., Leyland, P. C., Seal, R. L., Goodman, J. L., Thurmond, J., Strelets, V. B. and Wilson, R. J., and the FlyBase consortium. (2013) FlyBase: improvements to the bibliography. Nucleic Acids Res., 41, D751–D757

75. Costanzo, M. C., Engel, S. R., Wong, E. D., Lloyd, P., Karra, K., Chan, E. T., Weng, S., Paskov, K. M., Roe, G. R., Binkley, G., et al. (2014) Saccharomyces genome database provides new regulation data. Nucleic Acids Res., 42, D717–D725

76. Phanstiel, D. H., Brumbaugh, J., Wenger, C. D., Tian, S., Probasco, M. D., Bailey, D. J., Swaney, D. L., Tervo, M. A., Bolin, J. M., Ruotti, V., et al.(2011) Proteomic and phosphoproteomic comparison of human ES and iPS cells. Nat. Methods, 8, 821–827

77. Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T. Z., Garcia-Hernandez, M., Foerster, H., Li, D., Meyer, T., Muller, R., Ploetz, L., et al. (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Res., 36, D1009–D1014

78. Suzuki, A., Wakaguri, H., Yamashita, R., Kawano, S., Tsuchihara, K., Sugano, S., Suzuki, Y. and Nakai, K. (2015) DBTSS as an integrative platform for transcriptome, epigenome and genome sequence variation data. Nucleic Acids Res., 43, D87–D91

79. Sun, H., Wang, H., Zhu, R., Tang, K., Gong, Q., Cui, J., Cao, Z. and Liu, Q. (2014) iPEAP: integrating multiple omics and genetic data for pathway enrichment analysis. Bioinformatics, 30, 737–739

80. Bebek, G. and Yang, J. (2007) PathFinder: mining signal transduction pathway segments from protein-protein interaction networks. BMC Bioinformatics, 8, 335

81. Myers, C. L., Robson, D., Wible, A., Hibbs, M. A., Chiriac, C., Theesfeld, C. L., Dolinski, K. and Troyanskaya, O. G. (2005) Discovery of biological networks from diverse functional genomic data. Genome Biol., 6, R114

82. Ourfali, O., Shlomi, T., Ideker, T., Ruppin, E. and Sharan, R. (2007) SPINE: a framework for signaling-regulatory pathway inference from

cause-effect experiments. Bioinformatics, 23, i359–i366

83. Basha, O., Tirman, S., Eluk, A. and Yeger-Lotem, E. (2013) ResponseNet2.0: revealing signaling and regulatory pathways connecting your proteins and genes—now with human data. Nucleic Acids Res., 41, W198–W203

84. Lan, A., Smoly, I. Y., Rapaport, G., Lindquist, S., Fraenkel, E. and Yeger-Lotem, E. (2011) ResponseNet: revealing signaling and regulatory networks linking genetic and transcriptomic screening data. Nucleic Acids Res., 39, W424–W429

85. Wang, K., Saito, M., Bisikirska, B. C., Alvarez, M. J., Lim, W. K., Rajbhandari, P., Shen, Q., Nemenman, I., Basso, K., Margolin, A. A., et al. (2009) Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. Nat. Biotechnol., 27, 829–837

86. Zhu, F. and Guan, Y. (2014) Predicting dynamic signaling network response under unseen perturbations. Bioinformatics, 30, 2772–2778

87. Chen, J. and Zhang, S. (2016) Integrative analysis for identifying joint modular patterns of gene-expression and drug-response data. Bioinformatics, 32, 1724–1732

88. Qin, J., Li, M. J., Wang, P., Zhang, M. Q. and Wang, J. (2011) ChIP-Array: combinatory analysis of ChIP-seq/chip and microarray gene expression data to discover direct/indirect targets of a transcription factor. Nucleic Acids Res., 39, W430–W436

89. Wang, P., Qin, J., Qin, Y., Zhu, Y., Wang, L. Y., Li, M. J., Zhang, M. Q. and Wang, J. (2015) ChIP-Array 2: integrating multiple omics data to construct gene regulatory networks. Nucleic Acids Res., 43, W264–W269

90. Wang, S., Sun, H., Ma, J., Zang, C., Wang, C., Wang, J., Tang, Q., Meyer, C. A., Zhang, Y. and Liu, X. S. (2013) Target analysis by integration of transcriptome and ChIP-seq data with BETA. Nat. Protoc., 8, 2502–2515

91. Qin, J., Hu, Y., Xu, F., Yalamanchili, H. K. and Wang, J. (2014) Inferring gene regulatory networks by integrating ChIP-seq/chip and transcriptome data via LASSO-type regularization methods. Methods, 67, 294–303

92. Maienschein-Cline, M., Zhou, J., White, K. P., Sciammas, R. and Dinner, A. R. (2012) Discovering transcription factor regulatory targets using gene expression and binding data. Bioinformatics, 28, 206–213

93. Wu, G. and Ji, H. (2013) ChIPXpress: using publicly available gene expression data to improve ChIP-seq and ChIP-chip target gene ranking. BMC Bioinformatics, 14, 188

94. Tang, B., Hsu, H. K., Hsu, P. Y., Bonneville, R., Chen, S. S., Huang, T. H. and Jin, V. X. (2012) Hierarchical modularity in ERα transcriptional network is associated with distinct functions and implicates clinical outcomes. Sci. Rep., 2, 875

95. Yan, B., Li, H., Yang, X., Shao, J., Jang, M., Guan, D., Zou, S., Van Waes, C., Chen, Z. and Zhan, M. (2013) Unraveling regulatory programs for NF-kappaB, p53 and microRNAs in head and neck squamous cell carcinoma. PLoS One, 8, e73656

96. Pique-Regi, R., Degner, J. F., Pai, A. A., Gaffney, D. J., Gilad, Y. and Pritchard, J. K. (2011) Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Res., 21, 447–455

97. Huang, J. C., Babak, T., Corson, T. W., Chua, G., Khan, S., Gallie, B. L., Hughes, T. R., Blencowe, B. J., Frey, B. J. and Morris, Q. D. (2007) Using expression profiling data to identify human microRNA targets. Nat. Methods, 4, 1045–1049

98. Liang, Z., Zhou, H., He, Z., Zheng, H. and Wu, J. (2011) mirAct: a web tool for evaluating microRNA activity based on gene expression data. Nucleic Acids Res., 39, W139–144

99. Nam, S., Li, M., Choi, K., Balch, C., Kim, S. and Nephew, K. P. (2009) MicroRNA and mRNA integrated analysis (MMIA): a web tool for examining biological functions of microRNA expression. Nucleic Acids Res., 37, W356–362

100. Sales, G., Coppe, A., Bisognin, A., Biasiolo, M., Bortoluzzi, S. and Romualdi, C. (2010) MAGIA, a web-based tool for miRNA and Genes Integrated Analysis. Nucleic Acids Res., 38W352–W359

101. Qin, J., Li, M. J., Wang, P., Wong, N. S., Wong, M. P., Xia, Z., Tsao, G. S., Zhang, M. Q. and Wang, J. (2013) ProteoMirExpress: inferring microRNA and protein-centered regulatory networks from high-throughput proteomic and mRNA expression data. Mol. Cell. Proteomics, 12, 3379–3387

102. Wang, L.Y., Wang, P., Li, M.J., Qin, J., Wang, X., Zhang, M.Q. and Wang, J. (2011) EpiRegNet: constructing epigenetic regulatory network from high throughput gene expression data for humans. Epigenetics, 6, 1505–1512

103. Guan, D., Shao, J., Deng, Y., Wang, P., Zhao, Z., Liang, Y., Wang, J. and Yan, B. (2014) CMGRN: a web server for constructing multilevel gene regulatory networks using ChIP-seq and gene expression data. Bioinformatics, 30, 1190–1192

104. Huang, G. T., Athanassiou, C. and Benos, P. V. (2011) mirConnX: condition-specific mRNA-microRNA network integrator. Nucleic Acids Res., 39, W416–W423

105. Zhang, S., Li, Q., Liu, J. and Zhou, X. J. (2011) A novel computational framework for simultaneous integration of multiple types of genomic data to identify microRNA-gene regulatory modules. Bioinformatics, 27, i401–i409

106. Zhang, S., Liu, C. C., Li, W., Shen, H., Laird, P. W. and Zhou, X. J. (2012) Discovery of multi-dimensional modules by integrative analysis of cancer genomic data. Nucleic Acids Res., 40, 9379–9391

107. Li, W., Zhang, S., Liu, C. C. and Zhou, X. J. (2012) Identifying multi-layer gene regulatory modules from multi-dimensional genomic data. Bioinformatics, 28, 2458–2466

108. Guan, D., Shao, J., Zhao, Z., Wang, P., Qin, J., Deng, Y., Boheler, K. R., Wang, J. and Yan, B. (2014) PTHGRN: unraveling post-translational hierarchical gene regulatory networks using PPI, ChIP-seq and gene expression data. Nucleic Acids Res., 42, W130–136

109. Wu, D., Wang, D., Zhang, M. Q. and Gu, J. (2015) Fast dimension reduction and integrative clustering of multi-omics data using low-rank approximation: application to cancer molecular classification. BMC Genomics, 16, 1022

110. Oyama, M., Kozuka-Hata, H., Tasaki, S., Semba, K., Hattori, S., Sugano, S., Inoue, J. and Yamamoto, T. (2009) Temporal perturbation of tyrosine phosphoproteome dynamics reveals the system-wide regulatory networks. Mol. Cell. Proteomics, 8, 226–231

111. Bodenmiller, B., Wanka, S., Kraft, C., Urban, J., Campbell, D., Pedrioli, P. G., Gerrits, B., Picotti, P., Lam, H., Vitek, O., et al. (2010) Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. Sci. Signal., 3, rs4

112. Caldana, C., Fernie, A. R., Willmitzer, L. and Steinhauser, D. (2012) Unraveling retrograde signaling pathways: finding candidate signaling molecules via metabolomics and systems biology driven approaches. Front. Plant Sci., 3, 267

113. Zhu, J., Sova, P., Xu, Q., Dombek, K. M., Xu, E. Y., Vu, H., Tu, Z.,

Brem, R. B., Bumgarner, R. E. and Schadt, E. E. (2012) Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation. PLoS Biol., 10, e1001301

114. Jha, A. K., Huang, S. C., Sergushichev, A., Lampropoulou, V., Ivanova, Y., Loginicheva, E., Chmielewski, K., Stewart, K. M., Ashall, J., Everts, B., *et al.* (2015) Network integration of parallel metabolic and transcriptional data reveals metabolic modules that regulate macrophage polarization. Immunity, 42, 419–430

115. Marbach, D., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., Allison, K. R., The DREAM5 Consortium, Kellis, M., Collins, J. J. *et al.* (2012) Wisdom of crowds for robust gene network inference. Nat. Methods, 9, 796–804

116. Hu, Z., Killion, P. J. and Iyer, V. R. (2007) Genetic reconstruction of a functional transcriptional regulatory network. Nat. Genet., 39, 683–687

117. Song, L., Zhang, Z., Grasfeder, L. L., Boyle, A. P., Giresi, P. G., Lee, B. K., Sheffield, N. C., Gräf, S., Huss, M., Keefe, D., *et al.* (2011) Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome Res., 21, 1757–1767

118. Kelly, T. K., Liu, Y., Lay, F. D., Liang, G., Berman, B. P. and Jones, P. A. (2012) Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res., 22, 2497–2506

119. Natarajan, A., Yardimci, G. G., Sheffield, N. C., Crawford, G. E. and Ohler, U. (2012) Predicting cell-type-specific gene expression from regions of open chromatin. Genome Res., 22, 1711–1722

120. Lan, X., Witt, H., Katsumura, K., Ye, Z., Wang, Q., Bresnick, E. H., Farnham, P. J. and Jin, V. X. (2012) Integration of Hi-C and ChIP-seq data reveals distinct types of chromatin linkages. Nucleic Acids Res., 40, 7690–7704

121. Doench, J. G. and Sharp, P. A. (2004) Specificity of microRNA target selection in translational repression. Genes Dev., 18, 504–511

122. Chen, X. (2004) A microRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development. Science, 303, 2022–2025

123. Wightman, B., Ha, I. and Ruvkun, G. (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. Cell, 75, 855–862

124. Vasudevan, S., Tong, Y. and Steitz, J. A. (2007) Switching from repression to activation: microRNAs can up-regulate translation. Science, 318, 1931–1934

125. Vasudevan, S. and Steitz, J. A. (2007) AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. Cell, 128, 1105–1118

126. Chen, Y., Wang, Y., Xuan, Z., Chen, M. and Zhang, M. Q. (2016) *De novo* deciphering three-dimensional chromatin interaction and topological domains by wavelet transformation of epigenetic profiles. Nucleic Acids Res., 44, e106

127. Djekidel, M. N., Liang, Z., Wang, Q., Hu, Z., Li, G., Chen, Y. and Zhang, M. Q. (2015) 3CPET: finding co-factor complexes from ChIA-PET data using a hierarchical Dirichlet process. Genome Biol., 16, 288

128. Whalen, S., Truty, R. M. and Pollard, K. S. (2016) Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. Nat. Genet., 48, 488–496

129. Wang, Z., Zang, C., Rosenfeld, J. A., Schones, D. E., Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Peng, W., Zhang, M. Q., *et al.* (2008) Combinatorial patterns of histone acetylations and methyla-

130. tions in the human genome. Nat. Genet., 40, 897–903

130. Karlić, R., Chung, H. R., Lasserre, J., Vlahoviček, K. and Vingron, M. (2010) Histone modification levels are predictive for gene expression. Proc. Natl. Acad. Sci. USA, 107, 2926–2931

131. Zhu, Y., Sun, L., Chen, Z., Whitaker, J. W., Wang, T. and Wang, W. (2013) Predicting enhancer transcription and activity from chromatin modifications. Nucleic Acids Res., 41, 10032–10043

132. Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B. E., van Oudenaarden, A., *et al.* (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. USA, 106, 11667–11672

133. Knouf, E. C., Garg, K., Arroyo, J. D., Correa, Y., Sarkar, D., Parkin, R. K., Wurz, K., O'Briant, K. C., Godwin, A. K., Urban, N. D., *et al.* (2012) An integrative genomic approach identifies p73 and p63 as activators of miR-200 microRNA family transcription. Nucleic Acids Res., 40, 499–510

134. Ritchie, M. D., Holzinger, E. R., Li, R., Pendergrass, S. A. and Kim, D. (2015) Methods of integrating data to uncover genotype-phenotype interactions. Nat. Rev. Genet., 16, 85–97

135. Kristensen, V. N., Lingjærde, O. C., Russnes, H. G., Vollan, H. K., Frigessi, A. and Børresen-Dale, A. L. (2014) Principles and methods of integrative genomic analyses in cancer. Nat. Rev. Cancer, 14, 299–313

136. Marbach, D., Roy, S., Ay, F., Meyer, P. E., Candeias, R., Kahveci, T., Bristow, C. A. and Kellis, M. (2012) Predictive regulatory models in *Drosophila melanogaster* by integrative inference of transcriptional networks. Genome Res., 22, 1334–1349

137. Sintupisut, N., Liu, P. L. and Yeang, C. H. (2013) An integrative characterization of recurrent molecular aberrations in glioblastoma genomes. Nucleic Acids Res., 41, 8803–8821

138. Palsson, B. and Zengler, K. (2010) The challenges of integrating multi-omic data sets. Nat. Chem. Biol., 6, 787–789

139. Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., *et al.* (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. Cell, 122, 947–956

140. Marson, A., Levine, S. S., Cole, M. F., Frampton, G. M., Brambrink, T., Johnstone, S., Guenther, M. G., Johnston, W. K., Wernig, M., Newman, J., *et al.*(2008) Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell, 134, 521–533

141. Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., *et al.* (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature, 441, 349–353

142. Takahashi, K. and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell, 126, 663–676

143. Anokye-Danso, F., Trivedi, C. M., Juhr, D., Gupta, M., Cui, Z., Tian, Y., Zhang, Y., Yang, W., Gruber, P. J., Epstein, J. A., *et al.* (2011) Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. Cell Stem Cell, 8, 376–388

144. Wu, S. M. and Hochedlinger, K. (2011) Harnessing the potential of induced pluripotent stem cells for regenerative medicine. Nat. Cell Biol., 13, 497–505

145. Buganim, Y., Faddah, D. A. and Jaenisch, R. (2013) Mechanisms and

models of somatic cell reprogramming. Nat. Rev. Genet., 14, 427–439

146. Gifford, C. A., Ziller, M. J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A. K., Kelley, D. R., Shishkin, A. A., Issner, R., *et al.* (2013) Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell, 153, 1149–1163

147. Mohn, F., Weber, M., Rebhan, M., Roloff, T. C., Richter, J., Stadler, M. B., Bibel, M. and Schübeler, D. (2008) Lineage-specific polycomb targets and *de novo* DNA methylation define restriction and potential of neuronal progenitors. Mol. Cell, 30, 755–766

148. Tsankov, A. M., Gu, H., Akopian, V., Ziller, M. J., Donaghey, J., Amit, I., Gnirke, A. and Meissner, A. (2015) Transcription factor binding dynamics during human ES cell differentiation. Nature, 518, 344–349

149. Choukrallah, M. A., Song, S., Rolink, A. G., Burger, L. and Matthias, P. (2015) Enhancer repertoires are reshaped independently of early priming and heterochromatin dynamics during B cell differentiation. Nat. Commun., 6, 8324

150. Sancho-Martinez, I., Baek, S. H. and Izpisua Belmonte, J. C. (2012) Lineage conversion methodologies meet the reprogramming toolbox. Nat. Cell Biol., 14, 892–899

151. Heinäniemi, M., Nykter, M., Kramer, R., Wienecke-Baldacchino, A., Sinkkonen, L., Zhou, J. X., Kreisberg, R., Kauffman, S. A., Huang, S. and Shmulevich, I. (2013) Gene-pair expression signatures reveal lineage control. Nat. Methods, 10, 577–583

152. Cahan, P., Li, H., Morris, S. A., Lummertz da Rocha, E., Daley, G. Q. and Collins, J. J. (2014) CellNet: network biology applied to stem cell engineering. Cell, 158, 903–915

153. Rackham, O. J., Firas, J., Fang, H., Oates, M. E., Holmes, M. L., Knaupp, A. S., Suzuki, H., Nefzger, C. M., Daub, C. O., Shin, J. W., *et al.* (2016) A predictive computational framework for direct repro-gramming between human cell types. Nat. Genet., 48, 331–335

154. Li, M. J., Wang, P., Liu, X., Lim, E. L., Wang, Z., Yeager, M., Wong, M. P., Sham, P. C., Chanock, S. J. and Wang, J. (2012) GWASdb: a database for human genetic variants identified by genome-wide association studies. Nucleic Acids Res., 40, D1047–D1054

155. Li, M. J., Sham, P. C. and Wang, J. (2012) Genetic variant representation, annotation and prioritization in the post-GWAS era. Cell Res., 22, 1505–1508

156. Li, M. J., Yan, B., Sham, P. C. and Wang, J. (2015) Exploring the function of genetic variants in the non-coding genomic regions: approaches for identifying human regulatory variants affecting gene expression. Brief. Bioinform.16, 393–412

157. Alfaro, J. A., Sinha, A., Kislinger, T. and Boutros, P. C. (2014) Onco-proteogenomics: cancer proteomics joins forces with genomics. Nat. Methods, 11, 1107–1113

158. The Cancer Genome Atlas Research Network (2013) Integrated genomic characterization of endometrial carcinoma. Nature, 497, 67–73

159. Wang, B., Mezlini, A. M., Demir, F., Fiume, M., Tu, Z., Brudno, M., Haibe-Kains, B. and Goldenberg, A. (2014) Similarity network fusion for aggregating data types on a genomic scale. Nat. Methods, 11, 333–337

160. Yuan, Y., Van Allen, E. M., Omberg, L., Wagle, N., Amin-Mansour, A., Sokolov, A., Byers, L. A., Xu, Y., Hess, K. R., Diao, L., *et al.* (2014) Assessing the clinical utility of cancer genomic and proteomic data across tumor types. Nat. Biotechnol., 32, 644–652

161. Ernst, J. and Kellis, M. (2015) Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. Nat. Biotechnol., 33, 364–376

162. Marchini, J. and Howie, B. (2010) Genotype imputation for genome-wide association studies. Nat. Rev. Genet., 11, 499–511

163. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B. W., Jansen, R., de Geus, E. J., Boomsma, D. I., Wright, F. A., *et al.* (2016) Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet., 48, 245–252

164. Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. and Smyth, G. K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res., 43, e47

165. Orlando, D. A., Chen, M. W., Brown, V. E., Solanki, S., Choi, Y. J., Olson, E. R., Fritz, C. C., Bradner, J. E. and Guenther, M. G. (2014) Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. Cell Reports, 9, 1163–1170

166. Diaz, A., Park, K., Lim, D.A. and Song, J.S. (2012) Normalization, bias correction, and peak calling for ChIP-seq. Stat. Appl. Genet. Mol. Biol. 11, Article 9

167. Sysi-Aho, M., Katajamaa, M., Yetukuri, L. and Oresic, M. (2007) Normalization method for metabolomics data using optimal selection of multiple internal standards. BMC Bioinformatics, 8, 93

168. Johnson, W. E., Li, C. and Rabinovic, A. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics, 8, 118–127

169. Leek, J. T. and Storey, J. D. (2007) Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet., 3, e161

170. Thompson, J.A., Tan, J. and Greene, C.S. (2016) Cross-platform normalization of microarray and RNA-seq data for machine learning applications. PeerJ, 4, e1621

171. Zang, C., Wang, T., Deng, K., Li, B., Hu, S., Qin, Q., Xiao, T., Zhang, S., Meyer, C. A., He, H. H., *et al.* (2016) High-dimensional genomic data bias correction and data integration using MANCIE. Nat. Commun., 7, 11305

172. Rudy, J. and Valafar, F. (2011) Empirical comparison of cross-platform normalization methods for gene expression data. BMC Bioinformatics, 12, 467