Computational and Structural Biotechnology Journal 14 (2016) 1–7

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Mini Review

Blood transcriptomics and metabolomics for personalized medicine

Shuzhao Li [a,*], Andrei Todor [a], Ruiyan Luo [b]

[a] Department of Medicine, Emory University School of Medicine, 615 Michael Street, Atlanta, GA 30322, USA
[b] Division of Epidemiology and Biostatistics, School of Public Health, Georgia State University, One Park Place, Atlanta, GA 30303, USA

ARTICLE INFO

Article history:
Received 2 June 2015
Received in revised form 5 October 2015
Accepted 23 October 2015
Available online 31 October 2015

Keywords:
Transcriptomics
Metabolomics
Blood systems biology
Personalized medicine
Data integration

ABSTRACT

Molecular analysis of blood samples is pivotal to clinical diagnosis and has been intensively investigated since the rise of systems biology. Recent developments have opened new opportunities to utilize transcriptomics and metabolomics for personalized and precision medicine. Efforts from human immunology have infused into this area exquisite characterizations of subpopulations of blood cells. It is now possible to infer from blood transcriptomics, with fine accuracy, the contribution of immune activation and of cell subpopulations. In parallel, high-resolution mass spectrometry has brought revolutionary analytical capability, detecting >10,000 metabolites, together with environmental exposure, dietary intake, microbial activity, and pharmaceutical drugs. Thus, the re-examination of blood chemicals by metabolomics is in order. Transcriptomics and metabolomics can be integrated to provide a more comprehensive understanding of the human biological states. We will review these new data and methods and discuss how they can contribute to personalized medicine.

Contents

Many human diseases are complex and heterogeneous, whereas diagnostic methods are still limiting. Genetics and high-throughput molecular profiling now helps to redefine the disease classifications [1,2]. Personalized and precision medicine aims to design therapeutic interventions based on the condition of individual patients. For example, in the case of *trastuzumab*, a drug that is administered to breast cancer patients, its therapeutic efficiency varies depending on the patient's breast cancer subtype. This is because *trastuzumab* targets HER2 (human epidermal growth factor receptor type 2) proteins, and it is only effective on breast cancers with HER2 overexpression [3]. Therefore, a diagnostic test that determines HER2 overexpression is required before *trastuzumab* can be subscribed. A different type of example is adoptive T cell transfer for cancer immunotherapy, where specific T cells from an individual patient are engineered and expanded, then infused back to the same patient [4–6]. This type of therapy is "double" personalized because the T cells have to be from the very patient to be immunologically tolerant, and their surface receptors have to be specific to the tumor mutation found in that patient. Numerous examples exist that drug efficacy is limited due to the lack of "precision" mechanism. The widely used statins (cholesterol lowering drugs) may be efficacious in only 5% of the population, while esomeprazole (for heartburn treatment) fares even less [7]. A lot of research efforts have gone to identifying genetic variations associated with diseases, including many large genome-wide association studies (GWAS). However, genetic variations only account for small percentages of the occurrence of common

* Corresponding author.
E-mail address: shuzhao.li@gmail.com (S. Li).

http://dx.doi.org/10.1016/j.csbj.2015.10.005

diseases [8,9]. It is increasingly recognized that there is a large gap between genomics and phenotypes and that transcriptomics and metabolomics are important to fill this gap [10–14]. In this article, we will review the latest progress in transcriptomics and metabolomics, with a focus on samples from blood, a key tissue for clinical diagnosis. Since abundant introductory literature can be found on omics technologies and their data analysis, this article focuses more on important recent developments and opportunities.

1. An overdue review of "blood systems biology"

Blood has been intensively investigated since the beginning of molecular systems biology. Publications on disease diagnosis using blood transcriptomes are now numbered in thousands. Although it is widely recognized that mRNA only provides a slice of information from complex biology, few papers attempted to quantify the cell-level complexity in blood transcriptomics. Because blood is a mixture of many different cell types (Fig. 1), the fluctuation of cell populations alone causes large variations in transcriptomics data. This problem only became tractable with the recent progress in human immunology, where transcriptomics of isolated cell populations provided necessary information [15–17]. Nonetheless, a review on "blood systems biology" is long overdue.

As part of the body circulatory system, blood reflects the homeostasis of metabolism, hematopoietic development, and immune functions. As Fig. 1 shows, this involves many cell types and subtypes, and a number of "omics" technologies are employed to measure on different aspects of the system. The global molecular profiles of different cell types are tightly related to their developmental lineage and functions. As Novershtern et al. [18] showed, the clustering of transcriptomics data of blood cells reflects the hematopoietic process. The white blood cells are also sensitive indicators of the immune status. An infection will readily induce the influx of immune cells to blood as well as the activation of molecular programs in these cells. Cytokines and chemokines can increase dramatically during such events. The plasma contains molecular signals and wastes from the lymphatic system. The metabolites within plasma can reflect liver or kidney function, endocrine signaling, inflammation, and metabolic disorders. Thus, blood systems biology needs to address the following: (1) mixture data—most commonly, omics data are collected on peripheral blood mononuclear cells, where cell population composition is critical; (2) connection to a systemic model, such as pharmacokinetics or host-pathogen interaction models—blood is not a closed system by itself, only a window to systemic events; and (3) data integration. This could be the association between omics data and phenotype or the connection between different omics data types. We will start with an overview of transcriptomics and metabolomics then move on to specific topics for "blood systems biology".

2. Data acquisition of transcriptomics and metabolomics

DNA microarrays were developed in the 1990s as a major technology to measure transcriptomics. The technology relies on the specific hybridization between complementary polynucleotides. Probes are designed based on known gene transcripts and tethered on a glass surface. Targets are generated from biological samples, labeled directly or indirectly with fluorescent dyes. The hybridization reactions are carried on in miniaturized chambers. After the probes capture specific targets, the fluorescent signals are scanned and reported based on their grid locations. Thousands of microarray experiments are now deposited in public repositories such as GEO [19] and ArrayExpress [20].

As the cost of DNA sequencing drops, RNAseq becomes a viable alternative to capture transcriptomics. Using massively parallel sequencing platforms, RNAseq reads the number of DNA copies that are converted from mRNA, thus quantifying the concentration of mRNA species. From these sequencing reactions, the sequence variations in exons, such as single nucleotide polymorphisms (SNPs) and alternative splicing, are also captured in the data. Both the experimental methods and the computational analysis of RNAseq are evolving rapidly, and significant improvements are expected.

Metabolomics is the global profiling of small molecules (usually under 2000 Da). While nuclear magnetic resonance (NMR) [21] has been a powerful tool, mass spectrometry coupled with liquid or gas chromatography is the most popular platform due to the superior sensitivity and coverage [22–24]. The newest high-resolution mass spectrometer, in particular, yields unparalleled precision in analyzing chemicals in complex biological samples. The basic principle used by mass spectrometers is the differentiated deflection of charged particles in a magnetic field based on their mass. By the Lorentz law, the magnitude of the deflection is proportional to the mass to charge ratio. The advanced version, Fourier transform mass spectrometers, can achieve spectacular mass resolution by measuring the spinning frequency of ions that are trapped and oscillate in a chamber. The computational aspects of metabolomics are also in rapid progress, including open source feature extraction tools (XCMS [25], OpenMS [26], apLCMS [27], xMSanalyzer [28]), databases of metabolites (Human Metabolome

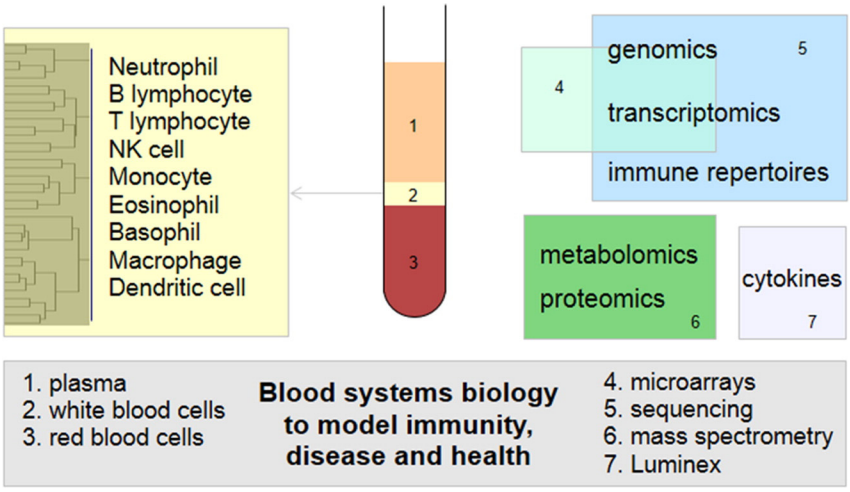

Fig. 1. Overview of blood systems biology, the pertinent samples and technologies. After a blood sample is taken, it is easily separated into plasma, white blood cells and red blood cells. The major white blood cells are listed on the left, while each cell type can be analyzed via exquisite protein markers via flow cytometry, giving information on particular subpopulations. Major "omics" technologies are listed on the right. DNA microarrays overlap with both genomics (genotyping arrays) and transcriptomics (expression arrays). DNA sequencing supports genomics (and epigenomics), transcriptomics (RNAseq), and immune repertoires. Immune repertoires include T cell receptor and B cell receptor sequences, whereas the latter represents antibody diversity. Both metabolomics (and environmental chemical exposures) and proteomics are largely dependent on mass spectrometry.