Published in final edited form as:

Trends Mol Med. 2018 January ; 24(1): 85–101. doi:10.1016/j.molmed.2017.11.001.

Panomics for Precision Medicine

Charanjit Sandhu^{1,*}, Alia Qureshi², and Andrew Emili¹

¹Donnelly Centre, University of Toronto, Toronto, Ontario, Canada

²Harvard Medical School, Beth Israel Deaconess Medical Center, Boston, MA USA

Abstract

Medicine is poised to undergo a digital transformation. High throughput platforms are creating terabytes of genomic, transcriptomic, proteomic and metabolomic data. The challenge is to interpret these data in a meaningful manner: to uncover relationships that are not readily apparent between molecular profiles and states of health or disease. This will require the development of novel data pipelines and computational tools. The combined analysis of multi-dimensional data is referred to as ‘panomics’. The ultimate hope of integrative panomics is to lead to the discovery and application of novel markers and targeted therapeutics that drive forward a new era of ‘Precision Medicine’ where inter-individual variation is accounted for in the treatment of patients.

Keywords

Precision Medicine; Genomics; Proteomics; Transcriptomics; Metabolomics; Panomics

The Big Data Era has Arrived

In clinical practice, patients present with a myriad of complaints. The clinician’s role is to interpret these complaints and perform a physical exam from which a differential diagnosis can be determined. Appropriate laboratory tests are then selected to rule in or out a given tentative diagnosis. Once the most likely diagnosis is selected, the patient is then started on a treatment plan. Each step in this process is frequently clouded by uncertainty. The patient’s presentation and laboratory findings are often inconsistent with each other and key pathognomonic features are often absent. The interpretation of a handful of laboratory investigations is used to make inferences on the functional status of a highly complex system consisting of approximately 35,000 genes giving rise to possibly six million protein variants¹². The limitations of this approach are reflected by post-test probabilities that are significantly less than one hundred percent³. As a result, the diagnosis is accepted to have

*Corresponding Authors - Contact Information: Charanjit Sandhu <charanjit.sandhu@mail.utoronto.ca>, (A.E.) DCCBR (room 914), 160 College St, Toronto, Ontario, Canada M5S 3E1, Phone: 416-946-7281; Fax: 416-978-8528; <andrew.emili@utoronto.ca>.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Resources

<https://www.cancer.gov/news-events/cancer-currents-blog/2017/tisagenlecleucel-fda-childhood-leukemia>

uncertainty and the patient's potential response to treatment is similarly variable. This variability in the response to treatment is reflected in a statistical measure referred to as the **number needed to treat (NNT)**: which represents the number of people who need to undergo treatment to prevent one adverse event⁴. Current practice would consider treating 10 patients to prevent one adverse event (NNT = 10) as efficacious therapy⁵. The hope and promise of '**panomics**' is to leverage large data sets of genetic and protein data to improve diagnostic accuracy and therapeutic efficacy.

The utilization of large high throughput DNA or protein data sets, so called 'Big Data' has been evolving over the last 15 years. Prior to the early 2000s, biochemical pathways had long been viewed and investigated as linear cascades. The completion of the Human Genome Project (HGP) ushered in the current era of holistic "pan genome" analysis⁶. This led to biochemical pathways being viewed as integrated networks and the development of a systems biology approach by biologists. **Next-generation DNA sequencing** technologies now allow for **whole genome "shotgun" sequencing** at a substantially reduced cost⁷. These technologies utilize short DNA reads that are subsequently assembled against DNA sequences generated by the HGP to produce individual genomes. Previously, unsequenced organisms required longer individual DNA reads and systematic mapping for sequence assembly. The advent of whole genome sequencing (WGS) platforms led to the rapid identification and fuller understanding of the sources of genetic variation: **chromosomal rearrangements, gene copy number variation, epigenetic changes and single nucleotide polymorphisms**. In parallel, developments in high throughput mass spectrometry platforms were similarly able to exploit the HGP to obtain in depth proteomic analysis of cells, biofluids and tissues. In this case, the HGP data has been used to generate a predicted in silico proteome against which peptide fragments identified in the mass spectrometer can subsequently be matched to assemble a proteomic expression profile.

Panomics is the integration of increasingly complex –'omics' data sets arising from the subfields of genomics, transcriptomics, proteomics and metabolomics -- to increase our understanding of the pathophysiology, diagnosis and treatment of disease (Figure 1). The future of medicine is one where panomics data are leveraged to create diagnostic tools with greater power and treatment, increasingly tailored to each patient's genetic variability. The ultimate goal will be the rapid diagnosis of disease, in a personalized approach, i.e. precision medicine. In order to realize this vision, significant challenges must be overcome. In this review, we offer an overview of the core issues and relevance of an integrative panomics platform that may enable the eventual success of precision medicine initiatives. We discuss the evolution of genomics and proteomics data, the targeted strategies currently employed to maximize the utility of the information generated and provide insights into areas where future performance gains might arise.

Genomics Usher in the Era of Big Data

Genomic research has focused primarily on cancer as it has long been understood to arise from the accumulation of multiple genetic alterations in genes that both drive cellular proliferation or arrest cellular growth (Box 1). One of the first cancers to undergo WGS was acute myeloid leukemia (AML), where ten mutations specific to tumor cells -- eight of

which were novel -- were identified⁸. The aim of characterizing such mutations was to identify molecular targets for therapy⁸.

Chronic myeloid leukemia (CML) was one of the first malignancies to have received targeted therapy⁹. Chromosomal translocation results in a constitutively active tyrosine-kinase (BCR-ABL) that is targeted for inhibition with tyrosine kinase inhibitors such as Imatinib. (although drug resistance and tumor recurrence remain important concerns)^{10,11}. Other prominent targeted therapies include the use of HER2/neu blockage in breast cancers with Herceptin¹², lymphoma kinase (ALK) inhibitors such as Crizotinib in lung cancer¹³, and EGFR receptor antagonists in lung and colon cancers, such as Gefitinib¹⁴. Despite this type of targeted approach, there are still some patients that do not respond to therapy or who present only a partial response even though the targeted mutation is present¹⁵. This likely reflects the presence of additional unscreened mutations in the biochemical network and reflects the limitations of targeting a single molecule. In the case of Imatinib resistance, point mutations in the BCR-ABL domain impaired binding of the drug¹⁵. Here, analyses of 'big data' may have the potential to generate hypotheses that are not readily apparent from individual sequence data, or from existing studies that have identified novel candidate target molecules. However, it is important to recognize that targeted approaches do not change the underlying genetic plasticity and heterogeneity that is a fundamental feature of tumors. As a consequence, results from clinical trials with targeted therapies have been less efficacious than initially hoped¹⁶. Recent advances in Adoptive Cell Transfer (ACT) technology, where T-cells are removed from patients and engineered to express chimeric T cell receptors targeting cancer cell antigens, have proven promising¹⁷. For instance, CAR-T cell ACT treatment for acute lymphoblast leukemia (ALL) has been recently approved by the FDA (Tisagenlecleucel; see resources). However, the efficacy of ACT therapy for various solid tumors remains to be established.

Cell free DNA (cfDNA) sequencing has recently been proposed to have a role in both the diagnosis and monitoring of disease recurrence¹⁸. In this case, cfDNA is recovered from body fluids including cerebrospinal fluid¹⁹, sputum²⁰, blood²¹, ascites²² or urine²³. In cancers, cfDNA is thought to be released as a result of tumor necrosis and apoptosis²⁴. To monitor disease recurrence, the mutation spectrum in cfDNA can be compared with previously identified mutations from the excised tumor²⁵. Profiling cfDNA has been proposed to monitor breast, lung and colorectal cancer occurrence, but larger studies are required for validation of preliminary results²¹. In lung cancers, Telomerase Reverse Transcriptase (hTERT) PCR amplification from cfDNA has been detected at a higher frequency in patients with non-small cell lung cancers when compared to disease free controls²⁶. Although extensive validation is required, the hope is that cfDNA may be able to one day complement information gleaned from other -omic platforms to improve the accuracy of existing diagnostic tests.

As the cost of WGS continues to drop, knowledge of a patient's DNA sequence may hopefully become a routine and critical variable used to direct test screening and to select therapy. In order to develop highly sensitive and specific tests, signal-to-noise challenges must be overcome, and this feat may be partly achieved with existing data. Here, big data analysis may be able to compensate for 'noise' by increasing the scale of data used for

analysis. Any ‘real’ signals represented by genetic variations in the form of **copy number variation** (CNV) or polymorphisms would be present across multiple data sets and be associated with a clinically significant phenotype²⁷. In order to achieve such big data collections, the speed, error rate and WGS cost must continue to improve. In the past decade, DNA sequencing improvements have exceeded the **Moore’s law equivalent** of computer power, by doubling every 24 months²⁸. The cost of sequencing the human genome has decreased from US\$10 million to below US\$1000 over the past decade (www.genome.gov).

Further performance gains might arise from ongoing improvements in microfluidics and assembly algorithms. For instance, **Next-generation sequencing** (NGS) platforms all currently use DNA templates either attached to a glass slide or to microbeads²⁹. The templates are amplified by PCR into clusters for sufficient optical detection before undergoing massively parallel sequencing reactions with fluorescently labeled nucleotides. The drawback of these platforms was initially the limited read length size and the need to develop novel alignment tools²⁹. Since their introduction, the Illumina NGS has become the platform with the lowest base per cost and the highest throughput with read lengths of up to 150 base pairs: sufficient for de novo genome assembly²⁹. Other improvements are ongoing, with an increasing number of transistors being added to microchips: DNA cluster densities have increased, optical detection has improved and newer enzymes have reduced sequencing time²⁹.

with existing platforms, Additional efficiency gains have arisen from a more targeted approach to genomic sequencing: **whole exome sequencing (WES)** versus WGS. The exomic DNA sequence (protein coding regions) comprises less than 2% of the human genome but contains the vast majority of known disease variants. The rationale for using WES is based on the idea that a targeted approach might dramatically reduce the sequencing requirements and other costs for conducting such studies. Accordingly, many clinical studies have focused on exomic sequencing although cost advantages are limited by the increased degree of sample preparation for exomic studies. The known exome variants have been compiled within the Exome Aggregation Consortium (ExAC)³⁰. Here, over three thousand human genes with complete truncation have been identified; of these, only 28% have been attributed an associated phenotype³⁰. For instance, recent studies have pooled exomic data with genotypic data, aiming to identify lipoprotein lipase mutations associated with coronary artery disease³¹ and exomic variants with an increased risk of stroke³². Some of the disadvantages of WES include an inability to detect structural variations, such as gene copy number, chromosomal translocations, inversions or deletions³³. Instead, WES may be optimal for identifying rare mutations with high penetrance and mendelian inheritance patterns. By contrast, diseases with complex genomic variation, including polymorphisms and structural variation may be best suited for WGS³³.

The next transformative change in DNA sequencing may occur with nanopore sequencing technologies that utilize single DNA molecules for sequencing in real time; a single DNA molecule is transited through a nanopore, and the DNA sequence can be inferred from transient electrical current changes³⁴. Molecules moving through nanopores for sequencing can consist of tagged nucleotides that produce unique ionic signatures from complementary

strand synthesis³⁵, free nucleotides released by exonuclease cleavage of DNA fragments³⁶ or native DNA fragments³⁷. Thus, the utilization of single molecule sequencing can eliminate the need for library amplification and with it, single cell genomics have the potential of enabling an improved understanding of tumor pathogenesis. Currently, these technologies require improvements to reduce the error rate relative to existing NGS platforms, which partly stems from the rapid DNA translocation through the nanopore, which can impair identification of the nucleotide with existing optical and electrical hardware³⁴. Nonetheless, fourth generation nanopore sequencing technologies hold the potential of reducing genome sequencing costs to below US \$1000³⁴.

Epigenetics and Personalized Medicine

Genetic variation includes acquired epigenetic changes that consists of modifications to the DNA template or chromatin, the nucleosomal packaged form of DNA that comprises each chromosome. These changes commonly consist of DNA methylation (–CH₃) and/or protein glycosylation, acetylation (–CH₃CO), phosphorylation or ubiquitination (Box 2).

Analysis of epigenetic changes is done using primarily two key methods: targeted chromatin immunoprecipitation (ChIP) or bisulfite sequencing of DNA. In the case of bisulfite modification (Figure 2), the addition of bisulfite to DNA results in the conversion of native cytosine to uracil, whereas methylated cytosines are protected. Upon subsequent Polymerase Chain Reaction (PCR) amplification of the DNA, the uracil is converted into thymidine. Microarrays with hybridization probes targeting CpG sites of interest (i.e. oligonucleotides complementary to both methylated and unmethylated sequences) can then aid in the identification of methylation states on a genome-scale level. Conversely, ChIP is performed by covalent stabilization of protein-DNA interactions by treatment with formaldehyde. The DNA-protein complexes are then extracted and sheared, and histone specific antibodies then added to enrich for specific protein-DNA components. After the covalent linkage between DNA and the bound protein is cleaved typically using proteases, the DNA is subsequently analyzed by PCR. ChIP assays can identify genome regions that are associated with specific histone modifications.

Ultimately, a patient's epigenome will be searched against a database of epigenomic maps for a spectrum of diseases to aid in patient diagnosis and management. The challenge is to first build this epigenomic database. Currently, several large-scale consortium projects are focused on creating such a database. For instance, The International Human Epigenomic Consortium (IHEC)³⁸, Deep Blue Epigenomic Data Server³⁹ and NIH Roadmap Epigenomics Mapping Consortium⁴⁰ contain maps of key site of histone modification and DNA methylation across hundreds of cell types and disease states. As the size of the database increases, big data analysis may provide a greater diagnostic accuracy and possibly pathognomic epigenetic profiles. For instance, In a recent study, autism spectrum disorders have been correlated with histone acetylation of common cis-regulatory DNA elements in the frontal and temporal cortex of human brains⁴¹. Ongoing studies are examining the impact of environmental exposures on gene expression patterns via epigenetic alterations. Occupational exposure to Bisphenol A, a chemical that is commonly used for the production of plastics has been associated with epigenetic modifications that can affect reproduction⁴².

In shift workers, DNA methylation in circadian genes has been noted to be increased⁴³ and in vitro exposure to carbon nanoparticles from non-scratchpaint has been associated with increased DNA methylation in an adenocarcinoma cell line⁴⁴.

Transcriptomics: From Microarrays to Single Cell RNA Sequencing

Genetic variation can manifest in the relative expression levels of RNA. NGS large scale transcriptional data that can be analyzed and integrated to identify hidden associations. These data have shown that non-coding (nc) RNA represents approximately three quarters of the transcribed human genome compared to less than two percent for the protein coding portion of the genome⁴⁵. The discovery of ncRNA has led to an increased interest in the role of regulatory RNA molecules as possible key mediators and biomarkers of disease (e.g. in cancer)⁴⁶. For example, microRNAs (miRNA) represent a subset of ncRNA molecules thought to regulate post transcriptional expression of more than half of all human genes⁴⁷. Due to the instability of mRNA in blood, biomarker studies have only recently focused primarily on mRNA levels recovered from cell extracts⁴⁸. Thus, the relatively recent search of ncRNAs, (e.g. miRNA) that are substantially more stable in blood than mRNAs, has opened new avenues of clinical investigation. Indeed, the increased resistance of ncRNAs to RNases in blood is thought to be conferred by lipid or protein carriers such as HDL, nucleophosmin 1 or the Ago2 protein⁴⁹. To date over 2500 putative human microRNAs have been reported in circulation⁵⁰.

Conversely, due to RNase susceptibility in plasma, studies of mRNA in blood have been limited to a handful of molecules showing increased stability. For instance, elevated levels of cyclin D1 mRNA in serum has been associated with a poorer prognosis and decreased response to tamoxifen chemotherapy in breast cancer patients with non-metastatic disease⁵¹. In addition, increased circulating TERT mRNA concentrations have been associated with a poor prognosis in patients with prostate cancer⁵², while preserved TGF-beta mRNA concentrations have been proposed as a possible biomarker for hepatocellular carcinoma⁵³.

Clinical studies of ncRNAs have for the most part assayed serum or blood miRNA levels⁵⁴. Future studies, might aim to examine other biologic fluids such as urine⁵⁵ or cerebral spinal fluid^{56,57} to reveal miRNA- or other ncRNA-based biomarkers, but rigorous testing will be required.. Accordingly, Long ncRNAs (lncRNA)⁵⁸ have been proposed to be of potential use in diagnosing chronic lymphocytic leukemias or multiple myeloma where LincRNA-21 and TUG1 were noted to be increased⁵⁹. Moreover, in hepatocellular carcinoma, elevated serum levels of HULC lncRNA⁶⁰ have been reported, while in oral cancer, a unique subset of lncRNA, MALALT-1 and HOTAIR, were found in saliva⁶¹. Yet, to our knowledge, urine long ncRNA (lncRNA) PCA3 is the only ncRNA currently used in a clinical setting, namely to assess disease progression in prostate cancer patients⁶².

Transcriptomic studies have now moved from using microarrays to NGS platforms. Microarray studies involve the isolation and tagging of RNA with fluorescent markers which are then hybridized to complementary strands tethered to a glass chip. The chips are scanned and fluorescent signals localized based on their grid location. The intensity of the fluorescent signal is correlated to the expression level of the specific RNA. NGS platforms

are likely to supersede DNA arrays as the economics of sequencing improve and the need to capture the vast repertoire of RNAs, including alternatively spliced transcripts, presents itself. Furthermore, specialized NGS platforms are facilitating single-cell RNA sequencing (scRNA-seq) which can enable a clearer understanding of cell identities, transition states, and functions when compared to bulk RNA samples⁶³. scRNA-seq also confers the ability to quantitate and identify RNA molecules specific to a certain cell population that may be critical in pathophysiology, including low copy RNA molecules that could potentially be discarded as noise in bulk RNA samples. Consequently, in the context of malignancies, scRNA might allow a clearer delineation between malignant and normal cells⁶⁴, and this appears to be the case in chronic myeloid leukemia (CML) where a distinct transcriptomal pattern was identified to persist during therapy, and which included genes previously implicated in CML, as well as novel candidate genes such as RAB31 (a GTPase), and SRSF2 (a spliceosome component)⁶⁴.

We anticipate that the coming decade might result in a substantial increase in transcriptomic and scRNA data, and be housed in the NCBI Gene Expression Omnibus database (GEO) and Sequence Read Archive database (SRA) databases, which are among the primary repositories of transcriptomic data. Nonetheless, challenges in the analysis of scRNA data include addressing the stochastic expression of genes in individual cells⁶⁵. In addition, given the rapidly growing accumulation and diversity of identified RNA sequences via RNAseq, further development of computational models to analyze, integrate and interpret such data will be evidently required before results from transcriptomic data can be potentially transferred to clinical settings.

The Unique Challenges and Promise of Proteomics

The evolution of precision medicine will require the integration of genomics with proteomic data. Proteomic analysis can provide a reliable measure of protein expression and the corresponding changes that occur in disease states^{66,67}. Since mRNA levels frequently do not correlate with protein abundance, due to regulatory factors such as ncRNA control, subcellular localization, post-translational modification and complex mechanisms dictating protein accumulation and destruction, direct protein measurements of protein abundance, interactions and modification states remain crucial. Given limitations with the specificity and affinity of **affinity capture reagents**, such as antibodies, protein mass spectrometry has thus emerged as a technology of choice for unbiased global analyses. The hope is that a protein network-centric approach to biomarker discovery may identify markers with increased sensitivity and specificity.

Early proteomic studies used 2-dimensional polyacrylamide gel electrophoresis: proteins migrate according to their mass and **isoelectric point** and are then eluted from the gel for subsequent proteolytic digestion and identification by mass spectrometry (MS). Here, the mass to charge ratio (m/z) of the parent peptide and so-called daughter fragments are searched against a spectral database of peptides for protein identification. The approach is informative, but tedious and biased towards detection of abundant components. Now, gel-free methods employing liquid chromatography coupled to tandem mass spectrometry have become the primary platform to measure protein mixtures rapidly on a proteome-wide scale

(Figure 3)⁶⁸. Blood fluid (plasma/serum, urine, ascites, cerebrospinal fluid, synovial fluid or saliva), cells or tissue are processed and the resulting peptide mixtures separated using one- or multi-dimensional liquid chromatography prior to detection and quantification by tandem mass spectrometry⁶⁹.

Acquisition of peptide information by tandem mass spectrometry may use data dependent, targeted or data independent strategies. The first method, data dependent acquisition (DDA) using LC-MS/MS is currently the most popular method for standard shotgun proteomics. In DDA, only peptides of the highest intensities in any given cycle are selected for fragmentation into daughter fragments; consequently, low-signal peptides pass through the instrument and remain unidentified. Multidimensional protein identification technology (MudPIT)^{68,70}, although relying on DDA, attempts to circumvent this issue by increasing peptide retention time to improve peptide separation. However, significant stochastic variation remains in the peptides identified in different samples. This can reduce the statistical power of studies examining quantitative differences between samples. Targeted peptide monitoring (TPM) attempts to address this issue by targeting only peptides with a specific m/z ratio for fragmentation⁷¹. The rationale behind this approach is that by reducing the number of competing ions, the peptide of interest is more likely to have a dominant intensity in a given cycle and consequently be selected for fragmentation using much shorter analysis times per sample. More recent approaches to this problem have been enabled by technological improvements in mass spectrometry instrumentation⁷², and include data-independent acquisition (DIA) where multiple peptides within a specific m/z range are cofragmented and the spectra of daughter fragments are measured simultaneously to capture a greater proportion of the peptides⁷³. Ongoing improvement in tandem mass spectrometry instrumentation, accuracy of spectral libraries⁷⁴ and peptide identification algorithms⁷⁵ are all key to increasing peptide capture and reducing the stochastic variation in peptide identification between samples.

The ultimate aim of many proteomic studies is to identify proteins that might be differentially expressed in a disease state. A t-test or other statistical methods, such as **gene set enrichment analysis** – once correction for multiple comparisons has been made -- is utilized to identify markers of interest. Statistically significant proteins of high interest would presumably be those linked to specific biochemical pathways implicated in a disease process. Functional annotations, such as **Gene Ontology (GO) bioprocess** terms, might then be utilized to infer the functional classes associated with proteins enriched in a pathophysiologic state⁷⁶. More recent approaches involve the construction of a network of differentially expressed proteins and their interrelationships, such as protein-protein interactions⁷⁷. Clusters within these networks can be examined to predict biochemical pathways likely showing altered activity in a given disease.

Recent studies have focused on differential expression of biochemical pathways as opposed to single markers of disease. For instance, exosomal protein (PSMA7), which is linked to proteasomal and inflammatory activity, was shown to be increased in the saliva of patients with inflammatory bowel disease (IBD)⁷⁸. In the case of gastric cancer, serum proteins involved in protein metabolic process have been found to be upregulated, while proteins involved in biological adhesion and developmental processes have been found to be

downregulated⁷⁹. In breast cancer cells, increased expression of proteins associated with protein synthesis and energy production have also been associated with disease progression⁸⁰. In hematological malignancies, differential phosphorylation of proteins have been reported to presumably distinguish between tumor cell types and patterns of kinase activity⁸¹.

While MS platforms are the mainstay of biomarker discovery, protein chips could be future tools for point of care analysis. Protein chips are glass slides containing immobilized probes such as antibodies or other affinity capture reagents such as MHC complexes of known specificity⁸². Proteins in serum or cell lysates are then hybridized to these chips to identify drug targets or protein-protein interactions. Thus, following rigorous testing and validation, it is possible that innovative chips targeting a panel of protein biomarkers might eventually provide rapid diagnostic results, particularly in combination with other assays, in a clinical setting, but robust technology is still far away from reaching this point.

Metabolomics

Metabolite levels are another indicator of gene activity and also reflect the effect of external influences on an organism. High-throughput metabolomics has recently emerged as a source of big data aimed at enabling precision medicine, and which will undoubtedly require open-access databases along with computational tools for data analysis. Furthermore, reference compounds are routinely used to generate libraries of spectral profiles that can be searched against unknown compounds that have undergone nuclear magnetic resonance (NMR) or MS analysis. Metabolomic data has already been reported to predict early drug toxicity in the form of hepatotoxicity or nephrotoxicity. For example, a decrease in the level in S-adenosylmethionine (SAME), a precursor of glutathione is a harbinger of increasing liver enzymes in rat models of liver damage following exposure to acetaminophen⁸³. Also, in rat models, exposure to cyclosporin A, a known nephrotoxic immunosuppressant, has been associated with elevated urine concentrations of trimethylamine and succinate⁸⁴. With significantly more validation, metabolomic data may help form a basis for future biomarkers in determining diagnosis and disease recurrence. The European Bioinformatics Institute (EMBL-EBI)-managed MetaboLights database and the NIH supported Metabolomics Workbench have been amongst the first databases for metabolomics⁸⁵⁸⁶. Metabolomic data pooled from these individual databases is found in the MetabolomeXchange⁸⁷. It is anticipated that progress in metabolomics will require ongoing data accumulation with consensus standards.

Currently, metabolomic assays can either be targeted or untargeted (global discovery). Targeted metabolomics refers to profiling a specific class of molecules such as lipids, amino acids or carbohydrates⁸⁸⁸⁹. This requires enrichment or **fractionation techniques** specific to the compound(s) being recovered. Untargeted metabolomics involves assaying all biomolecules detectable within a sample. Although MS coupled to liquid (or gas) chromatography has been used as a preferred method, currently, there are no metabolomic platforms that can efficiently assay all biomolecule classes in a single analysis; further, more than one analytic platform (such as Nuclear Magnetic Resonance, NMR), is typically used for metabolomic profiling. As with proteomics, component identification occurs by

matching MS spectra against a library of known or predicted fragment patterns, but this field is less developed than for proteomics.

Metabolomic analysis has shown that a person's diet, microbiome (gut flora), and age can influence their metabolome, the latter typified by metabolic changes, such as increased oxidative stress markers, with increasing age⁹⁰. Diagnostic studies have also shown specific metabolites associated with oral (putrescine and 8-hydroxyadenine) and pancreatic cancers (N-succinyl-L-diaminopimelic acid)⁹¹⁹². Metabolomic studies of *Mycobacterium sp.* have also identified twelve lipid markers including tuberculostearic acid that might be used as a tool to differentiate between pathogenic isolates⁹³ and to identify drug resistant strains⁹⁴⁹⁵. Furthermore, early identification of non-resistant bacteria in cases of active tuberculosis might be potentially exploited to reduce the burden of antibiotics, presumably leading to a reduction in the incidence of adverse events, such as for example retinopathy, peripheral neuropathy and hepatitis, which have been commonly associated with current treatments⁹⁶⁻⁹⁸. Fatty acid metabolomes in *Mycobacterium tuberculosis* strains have also been shown to predict microbial resistance to Rifampin⁹⁴. Specifically, a decrease in the concentrations of various 10-methyl branched-chain fatty acids and cell wall lipids has been associated with Rifampin resistance in these strains⁹⁴.

Pharmacometabolomics refers to examining changes in the metabolomic profile of a patient as a result of drug therapy. In the case of selective serotonin reuptake-inhibitors (SSRIs) which are used to treat mood related disorders such as depression and anxiety, a substantial proportion of patients fail to respond to treatment and have no substantial symptomatic improvement⁹⁹. In a cohort of patients with major depressive disorder (MDD) that were divided into responders and non-responders to citalopram (an SSRI), metabolomic studies have identified an increased level of glycine in the urine of non-responding patients⁹⁹. Subsequent genotyping of patients with MDD for genes encoding glycine synthesis and degradation enzymes led to the identification of a polymorphism in the glycine dehydrogenase (*GLDC*) gene that was associated with a poor response to treatment with SSRI⁹⁹. In this case, metabolomics led directly to the identification of a biomarker that could be used to identify non-responders to SSRIs and prevent months of futile treatment with drugs gradually titrated in increasing doses with the hope of finding a response to treatment⁹⁹.

Metabolomic studies have also identified markers for early drug toxicity. Hepatotoxicity secondary to amoxicillin-clavulanate in a clinical trial was associated with an early increase in the level of liver specific mi-RNA-122 and significant changes in urinary metabolites such as azelaic acid and 7-methylxanthine¹⁰⁰. Nephrotoxicity secondary to acyclovir in rats was associated with changes in metabolic pathways of arachidonic acid, tryptophan, arginine and proline, and glycerophospholipid¹⁰¹. In the case of toxicity to antibiotics such as vancomycin, tetracycline and metronidazole, the intestinal metabolome of mice detected approximately 87% of metabolites have an altered level to indicate a substantial disruption of intestinal homeostasis¹⁰². This includes metabolic pathways involving sugar, amino acid, bile acid and steroid hormone metabolism to dramatically change the chemical (and presumably microbial species) composition of feces¹⁰². In order to reduce the use of animals in toxicology studies, new strategies of using metabolomic studies of human cell cultures

and assessing changes in metabolic pathways may allow prioritization¹⁰³. A consortium NIH project for mapping the human toxome is attempting to create a database of pathways involved in the toxicological response¹⁰³.

Collectively, metabolomics studies might be used as complementary platform to substantiate other -omic technologies that aim to identify biomarkers for early diagnosis and prognosis of a number of pathologies. With ongoing improvements in mass spectrometry (and NMR) hardware, search algorithms and the accumulation of data sets, it may one day be possible to integrate metabolomic profiles with proteomics and transcriptomics datasets to contribute sensitive and specific diagnostic assays for many illnesses, but further testing is warranted.

The Healthcare Ecosystem and Precision Medicine

The development of Precision Medicine will require a closer integration of -omics and clinical data. For this to occur, a two-way communication between -omics and clinical databases will be required (Figure 4). Currently, disease specific databases compile disease specific variants with clinically significant outcomes. In the future, much larger population biobanks with associated clinical information might emerge that may accelerate the identification of relevant genetic variants. It is not clear if physician electronic medical record (EMR) platforms might be linked to these biobanks, given that obtaining corresponding clinical data remains a significant challenge. The infrastructure required for the development of precision medicine is currently absent. It is limited for many reasons, including the economics of WGS, the absence of population biobanks and existing privacy legislation.

As discussed, the underlying complexity of genetic variation may ultimately require large databases to generate studies with sufficient power to elucidate reliable predictors of disease. Big data analysis may be able to filter out the noise to more accurately identify biomarkers that are significant, comparing individuals with a given phenotype to those without. EMR platforms that communicate with biobanks in a two-way manner linking clinical outcomes with genome sequences may be able to accelerate the development of markers. For instance, standardized consent tools along with a road map for sharing genomic and clinical data has been proposed by the Regulatory and Ethics Working Group of the Global Alliance for Genomics and Health, but there are still many hurdles to overcome¹⁰⁴. Of note, a model for a future biobank has been provided by The Global Network of Personal Genome projects¹⁰⁵. Here, volunteers provide genomic data and corresponding health information so that phenotypes may be linked to specific genetic variants¹⁰⁶. In addition, standardization of EMR genetic data has also been proposed to promote the utilization of genomic data by the Displaying and Integrating Genetics Information through the EHR Action Collaborative (DIGITizE AC)¹⁰⁷. Nevertheless, the evolution of precision medicine evidently depends on the creation of improved infrastructure to capture and ethically distribute genomic data, as well as on the development of more powerful computational tools to be able to establish routine analyses.

Concluding Remarks

Big data is now being generated by high throughput multi-omics platforms with the hope of transforming health care by identifying novel targets and markers for diagnosis, prognosis and therapy. This will require ongoing improvements in hardware and software supporting the ‘panomics’ ecosystem. The optimization of nanopore technology for DNA sequencing may produce a substantial improvement in both the performance and cost over existing platforms, but this remains to be determined. The goal of sequencing individual genomes being at a cost below \$1000 USD might be achievable in the next five years given the well-known reduced rate of sequencing costs to date. This might also potentially promote a transition from exomic to WGS in clinical studies. Here, advances in algorithms that identify structural variations may be required to improve the characterization of gene copy number and other structural variations from WGS data.

In the case of proteomic and metabolomic data, a similar steady improvement in instrument technology and algorithms is required to reduce the stochastic variation associated with existing high throughput platforms. In addition, the use of increasingly automated platforms with improved in silico libraries harboring a greater inclusion of post-translational modifications may further our analysis, and consequently, enhance our understanding of disease pathogenesis. The integration of data from these panomics subfields will require ongoing improvements and standardization of data acquisition. In the case of neoplastic malignancies, this hope must be tempered with the realization that precision medicine cannot change the underlying genetic plasticity of tumors.

The challenges to achieving precision medicine are large but surmountable (see outstanding questions and Box 3). The cost of generating high-throughput data will need to decrease substantially as large data sets are required to filter out noise and identify relevant disease/phenotype associations. Furthermore, EMR platforms will need to integrate clinical databases with big data repositories of high-throughput data. Lastly, artificial intelligence applications might facilitate the mining of such databases. The ultimate outcome will hopefully satisfy the goals of precision medicine, with tangibly improved care and efficiency in the health care system.

Acknowledgments

AE acknowledges ongoing support as a Research Chair from the Ontario College of Universities, while CS did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Ewing B, Green P. Analysis of expressed sequence tags indicates 35,000 human genes. *Nat Genet.* 2000; 25(2):232–4. [published Online First: 2000/06/03]. DOI: 10.1038/76115 [PubMed: 10835644]
2. Ponomarenko EA, Poverennaya EV, Ilgisonis EV, et al. The Size of the Human Proteome: The Width and Depth. *International journal of analytical chemistry.* 2016; 2016:7436849. [published Online First: 2016/06/15]. doi: 10.1155/2016/7436849 [PubMed: 27298622]
3. Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet.* 2005; 365(9469): 1500–5. [published Online First: 2005/04/27]. DOI: 10.1016/s0140-6736(05)66422-7 [PubMed: 15850636]

4. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *Bmj*. 1995; 310(6977):452–4. [published Online First: 1995/02/18]. [PubMed: 7873954]
5. Furukawa TA. From effect size into number needed to treat. *Lancet*. 1999; 353(9165):1680. [published Online First: 1999/05/21]. doi: 10.1016/s0140-6736(99)01163-0
6. Green ED, Watson JD, Collins FS. Human Genome Project: Twenty-five years of big biology. *Nature*. 2015; 526(7571):29–31. [published Online First: 2015/10/04]. DOI: 10.1038/526029a [PubMed: 26432225]
7. Ke R, Mignardi M, Hauling T, et al. Fourth Generation of Next-Generation Sequencing Technologies: Promise and Consequences. *Hum Mutat*. 2016; [published Online First: 2016/07/14]. doi: 10.1002/humu.23051
8. Ley TJ, Mardis ER, Ding L, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature*. 2008; 456(7218):66–72. doi: nature07485 [pii] 10.1038/nature07485 [doi] [published Online First: 2008/11/07]. [PubMed: 18987736]
9. Ali MA. Chronic Myeloid Leukemia in the Era of Tyrosine Kinase Inhibitors: An Evolving Paradigm of Molecularly Targeted Therapy. *Molecular diagnosis & therapy*. 2016; 20(4):315–33. [published Online First: 2016/05/26]. DOI: 10.1007/s40291-016-0208-1 [PubMed: 27220498]
10. Yeung CC, Egan D, Radich JP. Molecular monitoring of chronic myeloid leukemia: present and future. *Expert review of molecular diagnostics*. 2016; :1–9. [published Online First: 2016/08/24]. DOI: 10.1080/14737159.2016.1227243
11. Pinilla-Ibarz J, Sweet KL, Corrales-Yepes GM, et al. Role of tyrosine-kinase inhibitors in myeloproliferative neoplasms: comparative lessons learned. *OncoTargets and therapy*. 2016; 9:4937–57. [published Online First: 2016/08/30]. DOI: 10.2147/ott.s102504 [PubMed: 27570458]
12. Baselga J, Coleman RE, Cortes J, et al. Advances in the management of HER2-positive early breast cancer. *Critical reviews in oncology/hematology*. 2017; 119:113–22. [published Online First: 2017/10/19]. DOI: 10.1016/j.critrevonc.2017.10.001 [PubMed: 29042085]
13. El-Osta H, Shackelford R. Personalized treatment options for ALK-positive metastatic non-small-cell lung cancer: potential role for Ceritinib. *Pharmacogenomics and personalized medicine*. 2015; 8:145–54. [published Online First: 2015/12/02]. DOI: 10.2147/pgpm.s71100 [PubMed: 26622190]
14. Yewale C, Baradia D, Vhora I, et al. Epidermal growth factor receptor targeting in cancer: a review of trends and strategies. *Biomaterials*. 2013; 34(34):8690–707. [published Online First: 2013/08/21]. DOI: 10.1016/j.biomaterials.2013.07.100 [PubMed: 23953842]
15. Shah NP, Nicoll JM, Nagar B, et al. Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. *Cancer cell*. 2002; 2(2):117–25. [published Online First: 2002/09/03]. [PubMed: 12204532]
16. Tannock IF, Hickman JA. Limits to Precision Cancer Medicine. *N Engl J Med*. 2017; 376(1):96–7. [published Online First: 2017/01/05]. DOI: 10.1056/NEJMc1613563
17. Schubert ML, Hoffmann JM, Dreger P, et al. Chimeric antigen receptor transduced T cells - Tuning up for the next generation. *Int J Cancer*. 2017; [published Online First: 2017/11/10]. doi: 10.1002/ijc.31147
18. Vandenbergh P, Wlodarska I, Tousseyn T, et al. Non-invasive detection of genomic imbalances in Hodgkin/Reed-Sternberg cells in early and advanced stage Hodgkin's lymphoma by sequencing of circulating cell-free DNA: a technical proof-of-principle study. *The Lancet Haematology*. 2015; 2(2):e55–65. [published Online First: 2015/12/22]. DOI: 10.1016/s2352-3026(14)00039-8 [PubMed: 26687610]
19. Pentsova EI, Shah RH, Tang J, et al. Evaluating Cancer of the Central Nervous System Through Next-Generation Sequencing of Cerebrospinal Fluid. *J Clin Oncol*. 2016; 34(20):2404–15. [published Online First: 2016/05/11]. DOI: 10.1200/jco.2016.66.6487 [PubMed: 27161972]
20. Su Y, Fang H, Jiang F. Integrating DNA methylation and microRNA biomarkers in sputum for lung cancer detection. *Clinical epigenetics*. 2016; 8:109. [published Online First: 2016/10/26]. doi: 10.1186/s13148-016-0275-5 [PubMed: 27777637]
21. Cree IA, Uttley L, Buckley Woods H, et al. The evidence base for circulating tumour DNA blood-based biomarkers for the early detection of cancer: a systematic mapping review. *BMC cancer*.

- 2017; 17(1):697. [published Online First: 2017/10/25]. doi: 10.1186/s12885-017-3693-7 [PubMed: 29061138]
22. Tokuhisa M, Ichikawa Y, Kosaka N, et al. Exosomal miRNAs from Peritoneum Lavage Fluid as Potential Prognostic Biomarkers of Peritoneal Metastasis in Gastric Cancer. *PLoS One*. 2015; 10(7):e0130472. [published Online First: 2015/07/25]. doi: 10.1371/journal.pone.0130472 [PubMed: 26208314]
 23. Foj L, Ferrer F, Serra M, et al. Exosomal and Non-Exosomal Urinary miRNAs in Prostate Cancer Detection and Prognosis. *Prostate*. 2017; 77(6):573–83. [published Online First: 2016/12/19]. DOI: 10.1002/pros.23295 [PubMed: 27990656]
 24. Henao Diaz E, Yachnin J, Gronberg H, et al. The In Vitro Stability of Circulating Tumour DNA. *PLoS One*. 2016; 11(12):e0168153. [published Online First: 2016/12/14]. doi: 10.1371/journal.pone.0168153 [PubMed: 27959945]
 25. Diaz LA Jr, Bardelli A. Liquid biopsies: genotyping circulating tumor DNA. *J Clin Oncol*. 2014; 32(6):579–86. [published Online First: 2014/01/23]. DOI: 10.1200/jco.2012.45.2011 [PubMed: 24449238]
 26. Coco S, Alama A, Vanni I, et al. Circulating Cell-Free DNA and Circulating Tumor Cells as Prognostic and Predictive Biomarkers in Advanced Non-Small Cell Lung Cancer Patients Treated with First-Line Chemotherapy. *Int J Mol Sci*. 2017; 18(5) [published Online First: 2017/05/12]. doi: 10.3390/ijms18051035
 27. Robasky K, Lewis NE, Church GM. The role of replicates for error mitigation in next-generation sequencing. *Nat Rev Genet*. 2014; 15(1):56–62. [published Online First: 2013/12/11]. DOI: 10.1038/nrg3655 [PubMed: 24322726]
 28. Agar J. The cybernetics moment: or why we call our age the information age / Moore's law: the life of Gordon Moore, silicon valley's quiet revolutionary. *Annals of science*. 2017; 74(1):88–90. [published Online First: 2016/09/15]. DOI: 10.1080/00033790.2016.1223345 [PubMed: 27624656]
 29. van Dijk EL, Auger H, Jaszczyszyn Y, et al. Ten years of next-generation sequencing technology. *Trends in genetics : TIG*. 2014; 30(9):418–26. [published Online First: 2014/08/12]. DOI: 10.1016/j.tig.2014.07.001 [PubMed: 25108476]
 30. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016; 536(7616):285–91. [published Online First: 2016/08/19]. DOI: 10.1038/nature19057 [PubMed: 27535533]
 31. Khera AV, Won HH, Peloso GM, et al. Association of Rare and Common Variation in the Lipoprotein Lipase Gene With Coronary Artery Disease. *Jama*. 2017; 317(9):937–46. [published Online First: 2017/03/08]. DOI: 10.1001/jama.2017.0972 [PubMed: 28267856]
 32. Auer PL, Nalls M, Meschia JF, et al. Rare and Coding Region Genetic Variants Associated With Risk of Ischemic Stroke: The NHLBI Exome Sequence Project. *JAMA neurology*. 2015; 72(7): 781–8. [published Online First: 2015/05/12]. DOI: 10.1001/jamaneurol.2015.0582 [PubMed: 25961151]
 33. Belkadi A, Bolze A, Itan Y, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A*. 2015; 112(17):5473–8. [published Online First: 2015/04/02]. DOI: 10.1073/pnas.1418631112 [PubMed: 25827230]
 34. Feng Y, Zhang Y, Ying C, et al. Nanopore-based fourth-generation DNA sequencing technology. *Genomics, proteomics & bioinformatics*. 2015; 13(1):4–16. [published Online First: 2015/03/07]. DOI: 10.1016/j.gpb.2015.01.009
 35. Fuller CW, Kumar S, Porel M, et al. Real-time single-molecule electronic DNA sequencing by synthesis using polymer-tagged nucleotides on a nanopore array. *Proc Natl Acad Sci U S A*. 2016; 113(19):5233–8. [published Online First: 2016/04/20]. DOI: 10.1073/pnas.1601782113 [PubMed: 27091962]
 36. Clarke J, Wu HC, Jayasinghe L, et al. Continuous base identification for single-molecule nanopore DNA sequencing. *Nature nanotechnology*. 2009; 4(4):265–70. [published Online First: 2009/04/08]. DOI: 10.1038/nnano.2009.12

37. Kasianowicz JJ, Brandin E, Branton D, et al. Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci U S A*. 1996; 93(24):13770–3. [published Online First: 1996/11/26]. [PubMed: 8943010]
38. Bujold D, Morais DA, Gauthier C, et al. The International Human Epigenome Consortium Data Portal. *Cell systems*. 2016; 3(5):496–99.e2. [published Online First: 2016/11/20]. DOI: 10.1016/j.cels.2016.10.019 [PubMed: 27863956]
39. Albrecht F, List M, Bock C, et al. DeepBlue epigenomic data server: programmatic data retrieval and analysis of epigenome region sets. *Nucleic Acids Res*. 2016; 44(W1):W581–6. [published Online First: 2016/04/17]. DOI: 10.1093/nar/gkw211 [PubMed: 27084938]
40. Bernstein BE, Stamatoyannopoulos JA, Costello JF, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol*. 2010; 28(10):1045–8. [published Online First: 2010/10/15]. DOI: 10.1038/nbt1010-1045 [PubMed: 20944595]
41. Sun W, Poschmann J, Cruz-Herrera Del Rosario R, et al. Histone Acetylome-wide Association Study of Autism Spectrum Disorder. *Cell*. 2016; 167(5):1385–97.e11. [published Online First: 2016/11/20]. DOI: 10.1016/j.cell.2016.10.031 [PubMed: 27863250]
42. Chianese R, Troisi J, Richards S, et al. Bisphenol A in reproduction: epigenetic effects. *Current medicinal chemistry*. 2017; [published Online First: 2017/10/11]. doi: 10.2174/0929867324666171009121001
43. Samulin Erdem J, Skare O, Petersen-Overleir M, et al. Mechanisms of Breast Cancer in Shift Workers: DNA Methylation in Five Core Circadian Genes in Nurses Working Night Shifts. *Journal of Cancer*. 2017; 8(15):2876–84. [published Online First: 2017/09/21]. DOI: 10.7150/jca.21064 [PubMed: 28928877]
44. Li J, Tian M, Cui L, et al. Low-dose carbon-based nanoparticle-induced effects in A549 lung cells determined by biospectroscopy are associated with increases in genomic methylation. *Scientific reports*. 2016; 6:20207. [published Online First: 2016/02/03]. doi: 10.1038/srep20207 [PubMed: 26831369]
45. Djebali S, Davis CA, Merkel A, et al. Landscape of transcription in human cells. *Nature*. 2012; 489(7414):101–8. [PubMed: 22955620]
46. Liu Y, Zong ZH, Guan X, et al. The role of long non-coding RNA PCA3 in epithelial ovarian carcinoma tumorigenesis and progression. *Gene*. 2017; [published Online First: 2017/09/03]. doi: 10.1016/j.gene.2017.08.027
47. Cho SH, An HJ, Kim KA, et al. Single nucleotide polymorphisms at miR-146a/196a2 and their primary ovarian insufficiency-related target gene regulation in granulosa cells. *PLoS One*. 2017; 12(8):e0183479. [published Online First: 2017/08/26]. doi: 10.1371/journal.pone.0183479 [PubMed: 28841705]
48. Danielson KM, Rubio R, Abderazzaq F, et al. High Throughput Sequencing of Extracellular RNA from Human Plasma. *PLoS One*. 2017; 12(1):e0164644. [published Online First: 2017/01/07]. doi: 10.1371/journal.pone.0164644 [PubMed: 28060806]
49. Wojciechowska A, Braniewska A, Kozar-Kaminska K. MicroRNA in cardiovascular biology and disease. *Advances in clinical and experimental medicine : official organ Wroclaw Medical University*. 2017; 26(5):865–74. [published Online First: 2017/10/27]. DOI: 10.17219/acem/62915 [PubMed: 29068585]
50. Zhao Y, Song Y, Yao L, et al. Circulating microRNAs: Promising Biomarkers Involved in Several Cancers and Other Diseases. *DNA and cell biology*. 2017; 36(2):77–94. [published Online First: 2017/01/07]. DOI: 10.1089/dna.2016.3426 [PubMed: 28060535]
51. Garcia V, Garcia JM, Pena C, et al. Free circulating mRNA in plasma from breast cancer patients and clinical outcome. *Cancer Lett*. 2008; 263(2):312–20. [published Online First: 2008/02/19]. DOI: 10.1016/j.canlet.2008.01.008 [PubMed: 18280643]
52. March-Villalba JA, Martinez-Jabaloyas JM, Herrero MJ, et al. Cell-free circulating plasma hTERT mRNA is a useful marker for prostate cancer diagnosis and is associated with poor prognosis tumor characteristics. *PLoS One*. 2012; 7(8):e43470. [PubMed: 22916267]
53. Dong ZZ, Yao DF, Yao M, et al. Clinical impact of plasma TGF-beta1 and circulating TGF-beta1 mRNA in diagnosis of hepatocellular carcinoma. *Hepatobiliary Pancreat Dis Int*. 2008; 7(3):288–95. [published Online First: 2008/06/05]. [PubMed: 18522884]

54. Shi T, Gao G, Cao Y. Long Noncoding RNAs as Novel Biomarkers Have a Promising Future in Cancer Diagnostics. *Disease markers*. 2016; 2016:9085195. [published Online First: 2016/05/05]. doi: 10.1155/2016/9085195 [PubMed: 27143813]
55. Huang X, Liang M, Dittmar R, et al. Extracellular microRNAs in urologic malignancies: chances and challenges. *Int J Mol Sci*. 2013; 14(7):14785–99. [PubMed: 23863690]
56. Machida A, Ohkubo T, Yokota T. Circulating microRNAs in the cerebrospinal fluid of patients with brain diseases. *Methods Mol Biol*. 2013; 1024:203–9. [published Online First: 2013/05/31]. DOI: 10.1007/978-1-62703-453-1_16 [PubMed: 23719953]
57. Teplyuk NM, Mollenhauer B, Gabriely G, et al. MicroRNAs in cerebrospinal fluid identify glioblastoma and metastatic brain cancers and reflect disease activity. *Neuro Oncol*. 2012; 14(6): 689–700. [PubMed: 22492962]
58. Xie C, Yuan J, Li H, et al. NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res*. 2014; 42(Database issue):D98–103. [PubMed: 24285305]
59. Isin M, Ozgur E, Cetin G, et al. Investigation of circulating lncRNAs in B-cell neoplasms. *Clin Chim Acta*. 2014; 431:255–9. [published Online First: 2014/03/04]. DOI: 10.1016/j.cca.2014.02.010 [PubMed: 24583225]
60. Xie H, Ma H, Zhou D. Plasma HULC as a promising novel biomarker for the detection of hepatocellular carcinoma. *Biomed Res Int*. 2013; 2013:136106. [PubMed: 23762823]
61. Tang H, Wu Z, Zhang J, et al. Salivary lncRNA as a potential marker for oral squamous cell carcinoma diagnosis. *Mol Med Rep*. 2013; 7(3):761–6. [published Online First: 2013/01/08]. DOI: 10.3892/mmr.2012.1254 [PubMed: 23292713]
62. Ronnau CG, Verhaegh GW, Luna-Velez MV, et al. Noncoding RNAs as novel biomarkers in prostate cancer. *Biomed Res Int*. 2014; 2014:591703. [PubMed: 25243154]
63. Cao J, Packer JS, Ramani V, et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*. 2017; 357(6352):661–67. [published Online First: 2017/08/19]. DOI: 10.1126/science.aam8940 [PubMed: 28818938]
64. Giustacchini A, Thongjuea S, Barkas N, et al. Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nat Med*. 2017; 23(6):692–702. [published Online First: 2017/05/16]. DOI: 10.1038/nm.4336 [PubMed: 28504724]
65. Kim JK, Marioni JC. Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data. *Genome Biol*. 2013; 14(1):R7. [published Online First: 2013/01/31]. doi: 10.1186/gb-2013-14-1-r7 [PubMed: 23360624]
66. Kuo KK, Kuo CJ, Chiu CY, et al. Quantitative Proteomic Analysis of Differentially Expressed Protein Profiles Involved in Pancreatic Ductal Adenocarcinoma. *Pancreas*. 2016; 45(1):71–83. [published Online First: 2015/08/12]. DOI: 10.1097/mpa.0000000000000388 [PubMed: 26262590]
67. Xiao H, Zhang Y, Kim Y, et al. Differential Proteomic Analysis of Human Saliva using Tandem Mass Tags Quantification for Gastric Cancer Detection. *Scientific reports*. 2016; 6:22165. [published Online First: 2016/02/26]. doi: 10.1038/srep22165 [PubMed: 26911362]
68. Washburn MP, Wolters D, Yates JR 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol*. 2001; 19(3):242–7. [doi] 85686 [pii] [published Online First: 2001/03/07]. DOI: 10.1038/85686 [PubMed: 11231557]
69. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*. 2003; 422(6928):198–207. [doi] nature01511 [pii] [published Online First: 2003/03/14]. DOI: 10.1038/nature01511 [PubMed: 12634793]
70. Link AJ, Eng J, Schieltz DM, et al. Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol*. 1999; 17(7):676–82. [published Online First: 1999/07/15]. DOI: 10.1038/10890 [PubMed: 10404161]
71. Sandhu C, Hewel JA, Badis G, et al. Evaluation of data-dependent versus targeted shotgun proteomic approaches for monitoring transcription factor expression in breast cancer. *J Proteome Res*. 2008; 7(4):1529–41. [doi] [published Online First: 2008/03/04]. DOI: 10.1021/pr700836q [PubMed: 18311902]
72. Gillet LC, Navarro P, Tate S, et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol*

- Cell Proteomics. 2012; 11(6):O111.016717. [published Online First: 2012/01/21]. doi: 10.1074/mcp.O111.016717
73. Weisbrod CR, Eng JK, Hoopmann MR, et al. Accurate peptide fragment mass analysis: multiplexed peptide identification and quantification. *J Proteome Res.* 2012; 11(3):1621–32. [published Online First: 2012/02/01]. DOI: 10.1021/pr2008175 [PubMed: 22288382]
74. Teleman J, Hauri S, Malmstrom J. Improvements in Mass Spectrometry Assay Library Generation for Targeted Proteomics. *J Proteome Res.* 2017; 16(7):2384–92. [published Online First: 2017/05/19]. DOI: 10.1021/acs.jproteome.6b00928 [PubMed: 28516777]
75. Tsou CC, Avtonomov D, Larsen B, et al. DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nature methods.* 2015; 12(3):258–64. 7 p following 64. [published Online First: 2015/01/20]. DOI: 10.1038/nmeth.3255 [PubMed: 25599550]
76. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005; 102(43):15545–50. [published Online First: 2005/10/04]. DOI: 10.1073/pnas.0506580102 [PubMed: 16199517]
77. Varghese RS, Zuo Y, Zhao Y, et al. Protein network construction using reverse phase protein array data. *Methods.* 2017; 124:89–99. [published Online First: 2017/06/28]. DOI: 10.1016/j.ymeth.2017.06.017 [PubMed: 28651964]
78. Zheng X, Chen F, Zhang Q, et al. Salivary exosomal PSMA7: a promising biomarker of inflammatory bowel disease. *Protein & cell.* 2017; [published Online First: 2017/05/20]. doi: 10.1007/s13238-017-0413-7
79. Yoo MW, Park J, Han HS, et al. Discovery of gastric cancer specific biomarkers by the application of serum proteomics. *Proteomics.* 2017; 17(6) [published Online First: 2017/01/31]. doi: 10.1002/pmic.201600332
80. Pozniak Y, Balint-Lahat N, Rudolph JD, et al. System-wide Clinical Proteomics of Breast Cancer Reveals Global Remodeling of Tissue Homeostasis. *Cell systems.* 2016; 2(3):172–84. [published Online First: 2016/05/03]. DOI: 10.1016/j.cels.2016.02.001 [PubMed: 27135363]
81. Casado P, Alcolea MP, Iorio F, et al. Phosphoproteomics data classify hematological cancer cell lines according to tumor type and sensitivity to kinase inhibitors. *Genome Biol.* 2013; 14(4):R37. [published Online First: 2013/05/01]. doi: 10.1186/gb-2013-14-4-r37 [PubMed: 23628362]
82. Hellstrom C, Dodig-Crnkovic T, Hong MG, et al. High-Density Serum/Plasma Reverse Phase Protein Arrays. *Methods Mol Biol.* 2017; 1619:229–38. [published Online First: 2017/07/05]. DOI: 10.1007/978-1-4939-7057-5_18 [PubMed: 28674890]
83. Schnackenberg LK, Chen M, Sun J, et al. Evaluations of the trans-sulfuration pathway in multiple liver toxicity studies. *Toxicology and applied pharmacology.* 2009; 235(1):25–32. [published Online First: 2008/12/24]. DOI: 10.1016/j.taap.2008.11.015 [PubMed: 19103213]
84. Lenz EM, Bright J, Knight R, et al. Cyclosporin A-induced changes in endogenous metabolites in rat urine: a metabonomic investigation using high field 1H NMR spectroscopy, HPLC-TOF/MS and chemometrics. *Journal of pharmaceutical and biomedical analysis.* 2004; 35(3):599–608. [published Online First: 2004/05/13]. DOI: 10.1016/j.jpba.2004.02.013 [PubMed: 15137985]
85. Kale NS, Haug K, Conesa P, et al. MetaboLights: An Open-Access Database Repository for Metabolomics Data. *Current protocols in bioinformatics.* 2016; 53:14.13.1–18. [published Online First: 2016/03/25]. DOI: 10.1002/0471250953.bi1413s53
86. Sud M, Fahy E, Cotter D, et al. Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res.* 2016; 44(D1):D463–70. [published Online First: 2015/10/16]. DOI: 10.1093/nar/gkv1042 [PubMed: 26467476]
87. Haug K, Salek RM, Steinbeck C. Global open data management in metabolomics. *Current opinion in chemical biology.* 2017; 36:58–63. [published Online First: 2017/01/17]. DOI: 10.1016/j.cbpa.2016.12.024 [PubMed: 28092796]
88. Zhou J, Yin Y. Strategies for large-scale targeted metabolomics quantification by liquid chromatography-mass spectrometry. *Analyst.* 2016; 141(23):6362–73. [published Online First: 2016/10/11]. DOI: 10.1039/c6an01753c [PubMed: 27722450]

89. Tugizimana F, Steenkamp PA, Pieter LA, et al. Mass Spectrometry in Untargeted LC-MS Metabolomics: ESI parameters and global coverage of the metabolome. *Rapid communications in mass spectrometry* : RCM. 2017; [published Online First: 2017/10/11]. doi: 10.1002/rcm.8010
90. Martinez-Gonzalez MA, Ruiz-Canela M, Hruby A, et al. Intervention Trials with the Mediterranean Diet in Cardiovascular Prevention: Understanding Potential Mechanisms through Metabolomic Profiling. *The Journal of nutrition*. 2016; [published Online First: 2016/03/11]. doi: 10.3945/jn.115.219147
91. Sridharan G, Ramani P, Patankar S. Serum metabolomics in oral leukoplakia and oral squamous cell carcinoma. *Journal of cancer research and therapeutics*. 2017; 13(3):556–61. [published Online First: 2017/09/02]. DOI: 10.4103/jcrt.JCRT_1233_16 [PubMed: 28862226]
92. He X, Zhong J, Wang S, et al. Serum metabolomics differentiating pancreatic cancer from new-onset diabetes. *Oncotarget*. 2017; 8(17):29116–24. [published Online First: 2017/04/19]. DOI: 10.18632/oncotarget.16249 [PubMed: 28418859]
93. Olivier I, Loots du T. A metabolomics approach to characterise and identify various *Mycobacterium* species. *Journal of microbiological methods*. 2012; 88(3):419–26. [published Online First: 2012/02/04]. DOI: 10.1016/j.mimet.2012.01.012 [PubMed: 22301369]
94. du Preez I, Loots du T. Altered fatty acid metabolism due to rifampicin-resistance conferring mutations in the *rpoB* Gene of *Mycobacterium tuberculosis*: mapping the potential of pharmacometabolomics for global health and personalized medicine. *Omics : a journal of integrative biology*. 2012; 16(11):596–603. [published Online First: 2012/09/13]. DOI: 10.1089/omi.2012.0028 [PubMed: 22966781]
95. Loots du T. An altered *Mycobacterium tuberculosis* metabolome induced by *katG* mutations resulting in isoniazid resistance. *Antimicrobial agents and chemotherapy*. 2014; 58(4):2144–9. [published Online First: 2014/01/29]. DOI: 10.1128/aac.02344-13 [PubMed: 24468786]
96. Ramappa V, Aithal GP. Hepatotoxicity Related to Anti-tuberculosis Drugs: Mechanisms and Management. *Journal of clinical and experimental hepatology*. 2013; 3(1):37–49. [published Online First: 2013/03/01]. DOI: 10.1016/j.jceh.2012.12.001 [PubMed: 25755470]
97. Kim KL, Park SP. Visual function test for early detection of ethambutol induced ocular toxicity at the subclinical level. *Cutaneous and ocular toxicology*. 2016; 35(3):228–32. [published Online First: 2015/09/13]. DOI: 10.3109/15569527.2015.1079784 [PubMed: 26361935]
98. Vistamehr S, Walsh TJ, Adelman RA. Ethambutol neuroretinopathy. *Seminars in ophthalmology*. 2007; 22(3):141–6. [published Online First: 2007/09/01]. DOI: 10.1080/08820530701457134 [PubMed: 17763233]
99. Ji Y, Hebring S, Zhu H, et al. Glycine and a glycine dehydrogenase (GLDC) SNP as citalopram/escitalopram response biomarkers in depression: pharmacometabolomics-informed pharmacogenomics. *Clinical pharmacology and therapeutics*. 2011; 89(1):97–104. [published Online First: 2010/11/26]. DOI: 10.1038/clpt.2010.250 [PubMed: 21107318]
100. Lee J, Ji SC, Kim B, et al. Exploration of Biomarkers for Amoxicillin/Clavulanate-Induced Liver Injury: Multi-Omics Approaches. *Clinical and translational science*. 2017; 10(3):163–71. [published Online First: 2016/10/28]. DOI: 10.1111/cts.12425 [PubMed: 27785887]
101. Zhang X, Li Y, Zhou H, et al. Plasma metabolic profiling analysis of nephrotoxicity induced by acyclovir using metabolomics coupled with multivariate data analysis. *Journal of pharmaceutical and biomedical analysis*. 2014; 97:151–6. [published Online First: 2014/05/28]. DOI: 10.1016/j.jpba.2014.04.036 [PubMed: 24863373]
102. Antunes LC, Han J, Ferreira RB, et al. Effect of antibiotic treatment on the intestinal metabolome. *Antimicrobial agents and chemotherapy*. 2011; 55(4):1494–503. [published Online First: 2011/02/02]. DOI: 10.1128/aac.01664-10 [PubMed: 21282433]
103. Bouhifd M, Hogberg HT, Kleensang A, et al. Mapping the human toxome by systems toxicology. *Basic & clinical pharmacology & toxicology*. 2014; 115(1):24–31. [published Online First: 2014/01/22]. DOI: 10.1111/bcpt.12198 [PubMed: 24443875]
104. Kosseim P, Dove ES, Baggaley C, et al. Building a data sharing model for global genomic research. *Genome Biol*. 15(8):430. doi: s13059-014-0430-2 [pii] 10.1186/s13059-014-0430-2 [doi] [published Online First: 2014/09/16].

105. Church GM. The personal genome project. *Molecular systems biology*. 2005; 1:2005.0030. [published Online First: 2006/05/27]. doi: 10.1038/msb4100040
106. Ball MP, Thakuria JV, Zaranek AW, et al. A public resource facilitating clinical use of genomes. *Proc Natl Acad Sci U S A*. 2012; 109(30):11920–7. [published Online First: 2012/07/17]. DOI: 10.1073/pnas.1201904109 [PubMed: 22797899]
107. Aronson SJ, Rehm HL. Building the foundation for genomics in precision medicine. *Nature*. 2015; 526(7573):336–42. [published Online First: 2015/10/16]. DOI: 10.1038/nature15816 [PubMed: 26469044]
108. Weinstein JN, Collisson EA, Mills GB, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 2013; 45(10):1113–20. [published Online First: 2013/09/28]. DOI: 10.1038/ng.2764 [PubMed: 24071849]
109. Ianov L, Riva A, Kumar A, et al. DNA Methylation of Synaptic Genes in the Prefrontal Cortex Is Associated with Aging and Age-Related Cognitive Impairment. *Frontiers in aging neuroscience*. 2017; 9:249. [published Online First: 2017/08/22]. doi: 10.3389/fnagi.2017.00249 [PubMed: 28824413]
110. Cheung WA, Shao X, Morin A, et al. Functional variation in allelic methylomes underscores a strong genetic contribution and reveals novel epigenetic alterations in the human epigenome. *Genome Biol*. 2017; 18(1):50. [published Online First: 2017/03/12]. doi: 10.1186/s13059-017-1173-7 [PubMed: 28283040]
111. Jacob S, Moley KH. Gametes and embryo epigenetic reprogramming affect developmental outcome: implication for assisted reproductive technologies. *Pediatr Res*. 2005; 58(3):437–46. doi: 58/3/437 [pii] 10.1203/01.PDR.0000179401.17161.D3 [doi] [published Online First: 2005/09/09]. [PubMed: 16148054]
112. Karami J, Mahmoudi M, Amirzargar A, et al. Promoter hypermethylation of BCL11B gene correlates with downregulation of gene transcription in ankylosing spondylitis patients. *Genes and immunity*. 2017; [published Online First: 2017/08/11]. doi: 10.1038/gene.2017.17
113. Gao SP, Sun HF, Li LD, et al. UHRF1 promotes breast cancer progression by suppressing KLF17 expression by hypermethylating its promoter. *American journal of cancer research*. 2017; 7(7): 1554–65. [published Online First: 2017/07/27]. [PubMed: 28744404]
114. Ngo TT, Yoo J, Dai Q, et al. Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nature communications*. 2016; 7:10813. [published Online First: 2016/02/26]. doi: 10.1038/ncomms10813
115. Grewal SI, Moazed D. Heterochromatin and epigenetic control of gene expression. *Science*. 2003; 301(5634):798–802. [published Online First: 2003/08/09]. DOI: 10.1126/science.1086887 [PubMed: 12907790]

Glossary

Affinity capture reagents

capture specific biomolecules such DNA, RNA, and proteins from a complex mixture. Most commonly utilizes beads with attached complementary nucleotides, antibodies or peptides to facilitate binding with the target biomolecule.

Big Data

extremely large data sets that are characterized by large volume, velocity and variety of data.

Cell free DNA (cfDNA) sequencing

sequencing of genomic material recovered from biological fluids that are free of intact cells. This includes plasma, urine, ascites, cerebrospinal fluid and pleural fluid.

Chromosomal rearrangements

changes in the native structure of a chromosome. This may include duplications, deletions, inversions and translocations.

Epigenetics

modifications to the DNA template or chromatin that may result in changes in gene expression. These changes commonly consist of DNA methylation (–CH₃) and/or protein acetylation (–CH₃CO), phosphorylation or ubiquitination.

Gene copy number variation

there are typically two copies of each gene within the genome. Gene copy number variation refers to the variability in the number of copies of a given gene; increased total expression of a gene is correlated with greater number of copies of a given gene.

Gene set enrichment analysis

computational method to identify genes or proteins enriched or depleted in a complex mixture.

Gene Ontology (GO) bioprocess

bioinformatics initiative to provide an organizational framework for biomolecules across species on the basis of their biological process, molecular function or cellular components.

Genomics

branch of molecular biology focused on DNA structure and function. The full DNA complement of DNA in a cell is referred to as the genome.

Fractionation techniques

separation of biomolecules on the basis of size, mass, charge or sequence.

High throughput

large scale data production platform that leverages automation.

In Silico proteome

the theoretical proteome and MS/MS daughter fragments that are predicted on the basis of the human genome project.

Isoelectric point

refers to the pH at which a biomolecule has no net charge.

Metabolomics

branch of molecular biology focused on cellular metabolites. The full complement of metabolites in each cell is referred to as the metabolome.

Moore's law

is the observation by Gordon Moore that the number of transistors in a dense integrated circuit doubles approximately every two years. This corresponds to a doubling of computing power every two years.

Next-generation DNA sequencing

high-throughput DNA sequencing platforms that sequence genomic data by performing complementary strand synthesis on a massively parallel scale. These short fragments are then aligned against a reference template or organized into overlapping fragments to assemble the genome.

Nuclear magnetic resonance (NMR)

application where an electromagnetic field is used to molecularly characterize biomolecules.

Number needed to treat (NNT)

represents the number of people who need to undergo treatment to prevent one adverse event.

Panomics

is the integration of –omics data sets arising from the subfields of genomics, transcriptomics, proteomics and metabolomics; aimed at increasing our understanding of biologic systems.

Precision Medicine

personalized approach to medicine where each patient's unique genetic composition is considered to guide diagnosis and management of the patient.

Proteomics

branch of molecular biology focused on protein composition. The full complement of protein in each cell is referred to as the proteome.

Single nucleotide polymorphisms (SNPs)

DNA sequence variation at a single nucleotide level.

Signal-to-noise ratio

refers to ratio of the strength of a true positive signal to a false positive signal.

Transcriptomics

branch of molecular biology focused on RNA composition. The full complement of RNA in each cell is referred to as the transcriptome.

Whole genome “shotgun” sequencing

sequencing of randomly fragmented DNA molecules and then reassembling the genome on the basis of overlapping regions or realignment against a reference template.

Box 1**Genomics Consortia**

Early genomic studies in the 1990s utilized microarrays to assess gene expression and genetic polymorphisms on the basis of hybridization between complementary polynucleotides tethered to a glass slide. Microarray studies were associated with an avalanche of data and were a precursor to the current era of big data produced by next generation DNA sequencing platforms.

The large data sets produced by DNA sequencing also outlined the need for a consortium approach to data discovery. The accumulation of cancer specific mutations with whole genome sequencing (WGS) led to the creation of The Cancer Genome Atlas (TCGA)¹⁰⁸, a large-scale international project to systematically identify cancer specific mutations in cancers with the highest incidence. Other largescale databases include The Gene Expression Omnibus database, a diverse repository of microarray and sequence based data and The Encyclopedia of DNA Elements (ENCODE) database which aims to organize the human genome into functional categories.

Box 2**Epigenetics Basics**

Epigenetic changes are primarily mediated by dedicated multi-component enzyme regulatory complexes, but oxidative reactants or nutritional metabolites can also catalyze epigenetic changes. The chemically altered genome is referred to as the epigenome and the net effects are changes in transcription or translation of DNA to either increase or decrease gene expression. Epigenetic changes have also been linked to aging, development and carcinogenesis¹⁰⁹¹¹⁰.

The most studied epigenetic change is DNA methylation which consists of the addition of a methyl group (–CH₃) to the cytosine nucleotide. This reaction is catalyzed by multiple classes of DNA methyltransferases (DNMTs). The pattern of DNA methylation is established early in development and thereafter remains relatively static within a given cell type¹¹¹. Regions that are frequently targeted for hypermethylation are known as CpG islands, which often represent conserved promoter regions enriched for cytosine and guanine in the DNA. The occurrence of repeating sequences of cytosine in the promoter region places these nucleotides in a key regulatory region of the DNA. Methylation of CpG dinucleotides in the promoter region has been typically linked to down regulation of gene expression; in ankylosing spondylitis patients, increased methylation of the *BCL11B* gene in peripheral blood mononuclear cells was associated with decreased levels of transcript¹¹². In breast cancer, increased expression of epigenetic regulator UHRF1, was associated with a poorer prognosis, thought to be mediated by increased methylation of the *KLF17* promoter, likely activated downstream of UHRF1¹¹³. These changes in gene expression have been thought to be secondary to the steric hindrance of DNA binding by transcription factors¹¹⁴.

Post translational modifications of histone proteins are also known to influence gene expression. DNA is packaged as chromatin by being wound tightly around nucleosomes, composed of various histone proteins. Epigenetic changes to histone proteins can consist of ubiquitination, phosphorylation of threonine or serine residues, or methylation and acetylation of arginine or lysine side chains. Chromatin remodeling can alter DNA tertiary or quaternary structure to either facilitate or downregulate gene expression. In contrast to epigenetic changes made to DNA, modification of histones appears to be a more dynamic process¹¹⁵. Genes may undergo periods of transcriptional activation or inactivation as a result of chromatin remodeling. For instance, acetylation of histones is typically associated with chromatin relaxation and subsequent gene activation, whereas histone methylation is associated with the formation of heterochromatin and the repression of gene expression¹¹⁵.

Box 3**Clinician's Corner**

Currently, clinical genomic studies focus on sequencing protein-coding exons. These studies are limited to identifying mutations with high penetrance and mendelian inheritance patterns. Ongoing technological improvements in DNA sequencing are required for a greater use of whole genome sequencing in clinical studies so that the role of structural variation can be assessed.

Cell free DNA sampled from biofluids might be potentially used as a future diagnostic and prognostic marker for certain cancers, but will require robust validation.

In the context of cancer, targeted therapies do not change the underlying genetic plasticity of tumors and as a consequence, the benefits of targeted therapy might be curtailed.

Unraveling the complexity of biological systems might be accelerated by improved machine learning algorithms that leverage phenotypic data with sequencing information.

Trends Box

- Genome sequencing costs are rapidly decreasing; within the coming decade, we might anticipate that whole genome sequencing may be more affordable for an increased number of patients.
- Automated high-throughput DNA sequencing and peptide sequencing platform are currently creating terabytes of information referred to as 'big data'.
- Big data are characterized by the three 'Vs': large volume of data, high velocity of data production occurring in real time, and the variety of data that can encompass multiple –omics subfields.
- The analysis of big data has the potential to identify novel biomarkers of disease and targets for therapy. The analysis of large scale data sets may enable the discovery of diagnostic or prognostic makers that are not readily apparent.
- The complexity and vastness of data analysis may ultimately require the development of computational platforms to aid in the discovery of biological pathways in association with health and disease.

Outstanding Questions Box

- Can cell-free DNA, recovered from biological fluids become a dominant diagnostic tool for certain cancers? The stability of DNA and rapid development of sequencing platforms suggests that this might be feasible in the future.
- What is the role and extent of specific non-coding RNA and epigenetic changes in driving specific disease processes? The accumulation of epigenetic changes and in some instances, subsequent gene silencing may have a significant role in the pathogenesis of many illnesses.
- Can proteomic platforms achieve their potential and identify peptides with low levels of expression in complex mixtures to function as effective diagnostic tools, or will tandem mass spectrometry platforms be primarily used to identify markers for protein-chip assays?
- Will metabolomics be able to consistently identify early drug toxicities in order to reduce possible adverse effects and presumably accelerate drug development?
- Can a future privacy framework be developed to establish a data pipeline between electronic medical records and biobanks to drive discovery and biomedical innovation?

Figure 1. Proposed Model of Precision Medicine Approaches

Data from -omics subfields are integrated (Panomics) to guide patient care in a manner that accounts for each patient's genetic variation.

Figure 2. Targeted Chromatin Immunoprecipitation

The technique involves covalent stabilization of protein-DNA interactions with formaldehyde. The DNA-protein complexes are immunoprecipitated and the DNA can be analyzed by microarray or whole genome sequencing.

Figure 3. Proteomic Analysis

Proteins undergo digestion with trypsin. Tryptic fragments are analysed by Tandem Mass Spectrometry (MS/MS). Mass to charge (m/z) ratios are determined for the parent and daughter fragments. Spectra are then matched with an in-silico database for protein identification.

Figure 4. Proposed Panomics Integration of Electronic Medical Records (EMR) with Biobanks using Artificial Intelligence for Biomarker Discovery

Data pipelines connecting EMR with repositories of –omics data may be required to facilitate big data analysis. Artificial intelligence algorithms may be able to search the multi-dimensional data to identify biomarker networks and targets for therapy relevant to disease types.