

# The Omics Revolution Continues: The Maturation of High-Throughput Biological Data Sources

Ewy Mathé<sup>1\*</sup>, John L. Hays<sup>2,3\*</sup>, Daniel G. Stover<sup>2\*</sup>, James L. Chen<sup>2,1</sup>

<sup>1</sup> Department of Biomedical Informatics, The Ohio State University, Columbus, OH, USA

<sup>2</sup> Department of Internal Medicine, The Ohio State University, Columbus, OH, USA

<sup>3</sup> Department of Obstetrics and Gynecology, The Ohio State University, Columbus, OH, USA

\* *Contributed Equally*

## Summary

**Objective:** The aim is to provide a comprehensive review of state-of-the-art omics approaches, including proteomics, metabolomics, cell-free DNA, and patient cohort matching algorithms in precision oncology.

**Methods:** In the past several years, the cancer informatics revolution has been the beneficiary of a data explosion. Different complementary omics technologies have begun coming into their own to provide a more nuanced view of the patient-tumor interaction beyond that of DNA alterations. A combined approach is beneficial to the patient as nearly all new cancer therapeutics are designed with an omics biomarker in mind. Proteomics and metabolomics provide us with a means of assaying in real-time the response of the tumor to treatment. Circulating cell-free DNA may allow us to better understand tumor heterogeneity and interactions with the host genome.

**Results:** Integration of increasingly available omics data increases our ability to segment patients into smaller and smaller cohorts, thereby prompting a shift in our thinking about how to use these omics data. With large repositories of patient omics-outcomes data being generated, patient cohort matching algorithms have become a dominant player.

**Conclusions:** The continued promise of precision oncology is to select patients who are most likely to benefit from treatment and to avoid toxicity for those who will not. The increased public availability of omics and outcomes data in patients, along with improved computational methods and resources, are making precision oncology a reality.

## Keywords

Circulating tumor DNA, proteomics, metabolomics, decision support techniques, algorithms, omics integration

Yearb Med Inform 2018;211-22

<http://dx.doi.org/10.1055/s-0038-1667085>

## The Continued Promise of Cancer Informatics

The future of cancer informatics is predicated on the continued development of methodologies that can identify key altered pathways that are susceptible to molecular targeted or immunologic therapies [1]. The increasing customization of medical treatment to specific patient characteristics has been possible through continued advances in a) our understanding of the physiologic mechanisms of disease through the proliferation of omics data (e.g., proteomics, metabolomics), and b) computing systems (e.g., patient matching algorithms) that facilitate matching and development of targeted agents [2]. These advancements allow for improved outcomes and for reduced exposure to the adverse effects of unnecessary treatment. They can help us better decipher the inter- (between patients) and intra- (different tumors within the same patient) tumor heterogeneity that is often a hurdle to treatment success and can contribute to both treatment failure and drug resistance [3]. Importantly, omics-based cancer medicine is here. In 2017, nearly 50% of the early-stage pipeline assets and 30% of late-stage molecular entities of pharmaceutical companies involved the use of biomarker tests [4]. Further, over a third of recent drug approvals have had DNA-based biomarkers included in their original US Food and Drug Administration (FDA) submissions [5]. We are also thinking about cancer informatics differently. Algorithmically, there has been

a shift from gene signatures to nonlinear approaches such as neural networks and advanced aggregative techniques to model complex relationships among patients [6]. Importantly, these approaches are the root of **cohort matching algorithms** that aim to find “patients like my patient.” Results of these algorithms are simpler to understand and have propelled the growth of clinical trials matching algorithms. National trials such as the NCI-MATCH [7], which pair patient tumors with specific tumor alterations to targeted medications, are a simplistic first step in this paradigm shift. The ability to perform complex matching, and matching rules, has relied on the growth of aggregated patient datasets and the ability to quickly assess tumor omics data.

This brief review focuses on three cancer omics data growth areas – proteomics, metabolomics, and circulating tumor and cell-free DNA. These omics approaches all try to enhance our current complex model of relationships among genes. We will also touch upon the paradigm shift from singular omics signatures to patient cohort matching – a shift that may potentially more readily take advantage of the large repositories of omics data that are being created. **Figure 1** underscores the foci of this review.

Within the past several years, tumor omics technologies have been integrating into clinical practice. Concurrently, we have increased our understanding of the underlying pathophysiology of not only the tumor, but also the patient/tumor interaction through this omics data. Acquisition of this omics

data, which is a focus of this review, has required improvements in detection techniques and data analysis. For example, assaying proteins using immunohistochemistry (IHC), the usage of singular antigens that bind to single proteins of interest in cancer tissue, is now being supplanted by mass spectrometry – which allows massively parallel identification of hundreds of proteins simultaneously. However, it has taken improved computer performance (and super computer clusters) to accurately identify this large number of proteins in a reasonable amount of time. This advancing field, **proteomics**, provides a far more accurate readout as compared to IHC – which is often subjective and difficult to parallelize.

Similarly, metabolite biomarkers have traditionally been singular molecules detected by immunoassay in the clinic. The chemotherapeutic drug methotrexate, for example, has levels that are detected via immunoassay for quantification purposes [8]. However, immunoassays only measure singular known metabolites and it is well known that combinations of metabolites are more clinically relevant than singular metabolites. With this in mind, **metabolomics** has emerged as a new omics field of study that aims to measure abundances of all small molecules detectable in biospecimens including blood, tissue, urine, and breath, among others. Typically, mass spectrometry (MS) and nuclear magnetic resonance (NMR) techniques are applied for measuring hundreds to thousands of metabolites in a given sample.

Advanced DNA sequencing, which ushered in the genomic revolution, has also improved greatly. Our ability to perform DNA sequencing with trace amounts of starting material (low-passage reads) with improved fidelity and detection is allowing us to detect circulating tumor DNA from the blood. **Circulating tumor DNA** (ctDNA) is tumor-derived fragmented DNA circulating in blood along with **cell free DNA** (cfDNA) from other sources (including normal cells) measuring about 150bp. ctDNA provides an overview of the genomic reservoir of different tumor clones and genomic diversity. ctDNA may finally provide a means of assaying intra-patient tumor heterogeneity – allowing us to get a sense of the relative abundance of genomic alterations across metastatic deposits

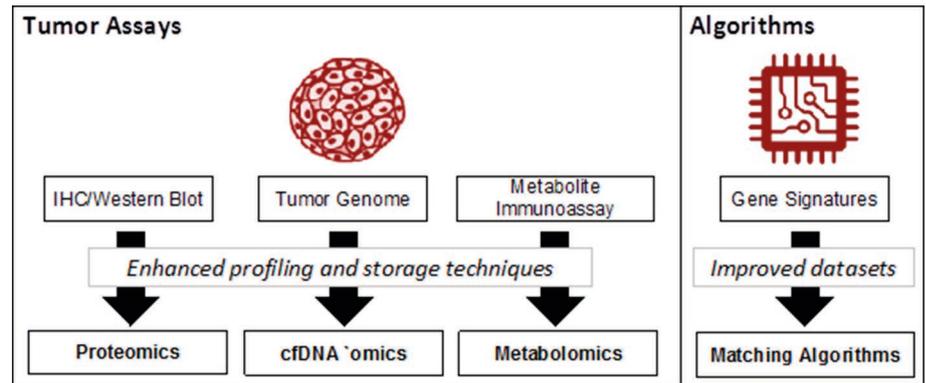


Fig. 1 Selected data and algorithm growth areas in cancer medicine.

within a patient. In the following sections, we will delve into each of these areas that have been introduced in the above.

## Proteomics

*Description of Technology:* The field of molecular therapeutics is a relatively novel approach that targets abnormalities in signaling pathways that play critical roles in tumor development and progression. While the genetic abnormalities of many conditions have been studied intensively, they do not always correlate with the phenotype of the disease. One possible explanation of this phenomenon is the lack of predictable changes in protein expression and function based solely on genetic information. One gene can encode for multiple proteins; protein concentration is temporally dynamic and protein compartmentalization is paramount to function; proteins are post translationally modified. All of this complexity leads to the importance of studying the proteome. Proteomics is fundamentally the study of proteins *and* their structure, functional status, localization, and interactions. This has only been possible as our understanding of proteins and their post-translational landscape has been investigated. Kinases and phosphatases that control the reversible process of phosphorylation and are dysregulated in many diseases including cancer have been studied individually for many years. However, only with the application of larger scale

technologies can we begin to understand the networks that control cellular phenotypes. Protein and lipid phosphorylation regulates cell survival signaling. Targeting kinases and phosphatases has proven to be paramount for improving therapeutic intervention in some diseases. In this regard, it is critical to define qualified cellular targets for cancer diagnosis and prognosis, as well as accurately predict and monitor responsiveness to therapies. Mutation profiling of selected genes or the whole exome can provide insights into possible activated pathways; however, to look at specific end points that can be targetable, one must examine the functional units of these mutational events, i.e., the protein.

*Recent Advancements:* There are multiple ways to examine events at the level of the protein. These range from Western blot level technologies which can examine a few proteins at a time, to mass spectrometry-based (MS-based) shotgun proteomics which can theoretically measure a very large subset, if not the entire, proteome. Broadly, most proteomic studies can be broken down into two categories: array-based and direct measurement. Array-based proteomic measurements are typically dependent on an antibody or substrate for a specific protein. Antibody-based proteomics platforms have been examined for the last 40 years and are still yielding exciting results. The most commonly used techniques for multiplexed assays are reverse phase protein arrays (RPPA), multiplexed immunofluorescence, and antibody-based chips/beads. These techniques

provide a quantitative assay that analyzes nanoliter amounts of sample for potentially hundreds of proteins simultaneously. These antibody-based assays determine quantitative levels of protein expression, as well as protein modifications such as phosphorylation, cleavage, and fatty acid modification [9-11]. Most techniques either array complex proteins samples and then probe with specific antibodies (e.g., RPPA), or array antibodies or specific ligands and then probe with a protein mixture. In essence, these assays have major strengths in identifying and validating cellular targets, characterizing signaling pathways and networks, as well as determining on and off target activity of novel drugs. One downside to array-based systems is the inherent reliance on quality antibodies or known substrates, which may or may not exist for all proteins of interest in a particular study. However, recent work has demonstrated a tissue-based map of the human proteome utilizing transcriptomic and multiplexed IHC-based techniques [12]. Similarly, The Cancer Protein Atlas (TCPA, <http://tcpaportal.org/tcpa/>) has examined samples collected during The Cancer Genome Atlas (TCGA) project and annotated selected samples with RPPA results. These initiatives provide a rich source of data at multiple levels from genes to transcripts to proteins [13].

Direct measurement techniques are based on the identification and quantification of the protein itself without utilizing analyte-based technologies that are solely dependent on the quality of the antibody. Most direct measurement techniques are based on MS approaches. MS-based proteomics techniques can be organized as bottom-up shotgun approaches which are able to accurately identify multiple proteins from complex mixtures. Complementary methods including stable isotope labeling (SILAC), tandem mass tags (TMT), and isobaric tags (iTRAQ) can be used in tandem with bottom-up approaches to measure relative or absolute concentrations of some or all proteins in complex mixtures. One of the inherent weaknesses of early MS-based approaches was the limited ability for absolute quantification of protein amounts. Given that many signaling events within cells are based upon changes in post-translational modifications with very small changes in to-

tal protein concentration, it was of particular interest to develop techniques that allowed for quantification of these changes. Perhaps one of the most exciting recent advances in MS-based proteomics techniques is the application of selected/multiple reaction monitoring (SRM/MRM) to quantify certain peptides of interest [14]. In contrast to array-based techniques described above, SRM-based methods can accurately measure multiple peptides from a single protein and theoretically measure multiple post-translational modifications simultaneously, independent of reliance on antibodies. In a study of multiple MS-based platforms, strong quantitative correlation to an immunoassay-based platform was observed for SRM using a synthetic peptide internal standard [15]. This contrasted to poor correlation for spectral counting, extracted ion chromatograms (XIC), and non-normalized SRM. The inherent flexibility of the various sectors associated with MS-based assay systems also allows for multiple questions to be asked that may not be feasible with antibody-based systems. For example, MS imaging can provide a molecular resolution of 2D tissue specimens [16]. This will allow for not only identification but also spatial relationships between biomarkers within samples. This enhanced level of information may be critical for defining pathway interactions or even more accurate molecular diagnostics.

*Clinical Utility:* Proteomics has a significant role to play in the translational analysis of patient tissue samples. The US National Institutes of Health (NIH) has recognized the importance of clinical proteomics with the establishment of the Office of Clinical Cancer Proteomics Research (OCCPR). One of the largest working groups established through the OCCPR is the Clinical Proteomic Tumor Analysis Consortium (CPTAC). Vasaikar et al., with support from the CPTAC, recently published a new data portal that links TCGA clinical genomics data with MS-derived proteomics data in a similar fashion to the work performed by the TCPA utilizing RPPA arrays [13, 17]. Of note, the CPTAC initiative has produced a number of publications [18-22] and freely available datasets for use by the broader omics community. As an example, Mundt et

al. have recently published an MS-based proteomics study on patient-derived xenografts to identify potential mechanisms of intrinsic and adaptive resistance to phosphoinositide 3-kinase inhibitors that will likely have clinical impact in the near future [21]. Much of the clinical utility of proteomics research will be driven by sample availability and quantity with accurate links to clinical data. While RPPA analysis requires very small amounts of sample, there is also a limited proteome sample space to test. MS-based techniques can test for a wide sample space of the proteome, but this requires a sample size that can hinder research. More recent MS techniques that are focused on quantitative analysis of a subset of the proteome can be done on smaller sample sizes and likely represent the future of MS-based clinical tests.

*Data Challenges:* Proteomics is a relatively new field in the world of large scale omics datasets. Older technologies that are array-based have relative straightforward datasets. Most RPPA datasets are reported in normalized linear or Log<sub>2</sub> median centered scales based on the detection ranges of the specific equipment being utilized. There is minimal data manipulation outside of identifying the linear range of each sample being measured through the use of internal standards and then extrapolating absolute quantification through interpolation of a standard curve [23]. However, with the explosion of MS-based proteomic techniques, cross platform data analysis and sharing have been associated with their fair share of growing pains. The Human Proteome Organization (HUPO) has taken the lead in defining requirements for proteomic submission and repositories. Such tasks that our colleagues in the genomics world have taken for granted over the past 15 years are now being reinvented for the proteomics field. The inherent complexity of proteomics data and the multiple platforms utilized make sharing data a non-trivial affair. There is also an issue of technology outpacing our reporting ability. While peptide and spectral libraries have been and continue to be important for most MS proteomic analysis and deposition (PRIDE and PeptideAtlas being the major resources) [24-26], there is also a need for a common library of

molecular transitions with the explosion of SRM/MRM techniques. PASSEL has been and continues to be the leading resource for SRM transition datasets [27]. Probably the most important advance in dataset submission and dissemination has been the continued development of the proteome exchange (PX) consortium. Beginning in 2011, the PX consortium has continually added on members to allow common data formats for all proteomic datasets. This will allow remarkable opportunities for data reanalysis and reinterpretation that our genomics colleagues have been enjoying for more than 10 years.

*Future Directions:* As more data from multiple tumor types become available, the ability to link genome to proteome, and ultimately phenotype and treatment choices, becomes less of a holy grail and more a clinical reality. The integrated information will display the potential therapeutic targets or biomarkers to accurately predict or rapidly define intracellular signaling networks and functional outcomes affected by therapeutics. Clinically, we are starting to see this through large basket-type trials incorporating genomic data matched to targeted drugs, independent of tumor type. The understanding of the proteomic context of a genomic alteration will be key to expanding the repertoire of successful biomarker-driven clinical trials. RPPA- and MS-based phosphoproteome investigation is already being explored in the context of pathway activation and targeted therapies. Similarly, utilizing targeted genomic mutation panels identifies a subset of ovarian cancer patients that may be sensitive to poly ADP ribose polymerase (PARP) inhibition, but incorporating proteomic analysis can also help identify possible responders in genomically unselected populations treated with cytotoxic chemotherapy and/or PARP inhibitors [28-31].

## Metabolomics

*Description of Technology:* The overall aim in metabolomics studies is to measure levels of small molecules, less than 1,500 Daltons, in a given biospecimen (e.g., blood, tissue,

urine, breath). The combination of various extraction (e.g., enrichment of lipids or protein-bound metabolites) and analytical techniques generates metabolic profiles that span many known and unknown metabolic pathways. Such metabolic profiles are a rich resource for defining phenotypes of distinct diseases such as cancer, and they reflect alterations in the genome, epigenome, proteome, and environment (exposures and lifestyle). For this reason, metabolomics is increasingly applied to complement other omics characterization of cells and clinical samples [32, 33], and is invaluable for uncovering putative clinical biomarkers, therapeutic targets, and aberrant biological mechanisms and pathways that are associated with cancer [34-40].

Metabolites can be categorized as endogenous, naturally produced by the host or cells under study, or exogenous, including drugs, foods, and cosmetics among others. While the goal is to measure all metabolites in a given biospecimen, analogous to measuring all gene levels in transcriptomic studies, current analytical acquisition techniques can only capture a fraction of metabolites given one assay or platform [36, 41]. For example, as of April 2018, the Human Metabolome Database [41-43] contains information on 114,100 metabolites, yet only 22,287 (19.5%) have been detected in human biospecimens. Also, unlike genomics and transcriptomics where one can measure genome-wide features (e.g., expression, variants) with one assay, metabolomics requires multiple analytical techniques and instrumentation for a broad coverage of metabolites (e.g., polar and nonpolar metabolites). In practice, a specific combination of sample preparation (e.g., enrichment of nonpolar metabolites) and analytical technique is often optimized for a certain class of compounds (e.g., lipids) [36].

The two main analytical approaches for measuring metabolites are NMR and MS [44-47]. Abundance detection by MS is typically preceded by a molecule separation technique such as liquid (LC) or gas (GC) chromatography. While NMR is considered the gold standard for compound identification (when analyzing singular compounds in pure form) and produces quantitative measures, MS-based methods are more

sensitive (e.g., able to detect low abundance metabolites) and detect more (e.g., several hundreds to thousands) metabolites [48].

Of note, metabolomics studies can be classified as targeted [49] or untargeted [50]. In targeted studies, a small (~1-150) panel of metabolites with known chemical characteristics and annotations are measured and the sample preparation and analytical platform used are optimized to minimize experimental artifacts. Examples of artifacts are fragmentation and adduct formation (e.g., addition of sodium or hydrogen ions) in electrospray ionization [51]. Measurements can be performed using standards and thus produce quantitative or semi-quantitative measurements. In contrast, untargeted metabolomics aims to detect all possible metabolites given a biospecimen. Untargeted approaches yield relative measurements of thousands of signals that represent known metabolites, experimental artifacts (e.g., adducts), or unidentified metabolites [52]. While many more metabolites can be captured with untargeted approaches, it is very challenging to annotate signals and identify metabolites [51]. Verification of metabolite identity requires prediction of elemental composition from accurate masses, and eventually, further experimentation (NMR being the gold standard) that requires the use of a purified standard for the metabolite of interest [45, 53-56]. If a purified standard is not commercially available, one must be synthesized in-house and thus this validation process could take several years. Ultimately, a targeted approach is favorable if there is *a priori* knowledge on the biological system or disease under study because measurements are quantitative and the data quality is high [52]. However, despite the high level of noise and the increased complexity in data analysis, untargeted approaches are favorable for discovering novel biomarkers or generating data-driven hypothesis [52].

*Recent Advancements:* As metabolomics strategies are increasingly being applied in biomedical research, advances in automation and improved quantification of NMR- and MS-based methods are producing high throughput, reproducible data [44, 57, 58]. Integration of NMR and LC-MS techniques is increasingly applied to enhance repro-

ducibility, metabolite identification, and to ensure measurement integrity [59]. Such improvements in data acquisition techniques are critical for expanding the coverage of metabolites that can be reliably measured. At the same time, these advances are producing larger data, requiring the construction of databases and the development of data analysis methods, tools, and pipelines. Currently, the two major sources of publicly available data are the Metabolomics Workbench [60] and MetaboLights [61]. The Metabolomics Workbench, sponsored by the NIH Common Fund, also provides access to analytical protocols (e.g., sample preparation and analysis), metabolite standards, computational tools, and training. While metabolomics data is very informative, for example, to uncover putative clinical biomarkers, understanding how metabolites are produced and their function further deepens our understanding of disease phenotypes and mechanisms. In turn, this mechanistic understanding can guide the search for putative drug targets. With this in mind, integration of metabolomics data with other omics datasets, including genome, proteome, and microbiome, is increasingly performed [62, 63]. Integration of omics datasets includes numerical integration techniques such as canonical correlations or multivariate modeling, and network/pathway based approaches [64-69]. Furthermore, open-source user-friendly software for metabolomics analysis and interpretation through pathway analysis has been critical for guiding analysis and interpretation of the data. Examples include XCMS [70, 71], MetaboAnalyst [72, 73], and Metabox [74].

**Data Challenges:** Unlike genomics, a reference metabolome does not exist and it is currently impossible to measure all metabolites in a given biospecimen. This lack of reference causes many data analysis issues, particularly for untargeted metabolomics studies where the identification of metabolites is difficult to pin down [75]. The field also suffers from multiple metabolite naming conventions. In fact, different naming conventions are more appropriate for certain types of data acquisition techniques. For example, while untargeted metabolomics approaches cannot resolve differing stereochemistry or double bond position/

geometry, other approaches can identify metabolites with more or less granularity [60]. Translation services, including Refmet [60] and the Chemical Translation Service [76] help in that regard. Also, the multitude of data acquisition techniques makes it difficult to organize the data in a standardized fashion [77]; instrumentation vendors have specific data formats that are tied to proprietary software and conversion of these file formats to open source formats can require specific operating systems or software licenses. Differences in how the data was generated also make it difficult to compare results across studies. With many missing identities and different resolution of identification, it is difficult to map a metabolite from one study to another. Standardization is thus critical to handle such challenges but is in their nascence [60, 77]. Standard protocols for downstream data analyses, including quality control, transformation/normalization, and differential analysis, are also difficult to establish, namely due to differences in experimental study design and data acquisition. Although publicly available tools and software aim to provide standard approaches [70-74, 78], detailed descriptions of parameters (e.g., mass divided by charge number [m/z] range allowed for binning features) and cutoffs used are often lacking in published work and makes reproducibility of results difficult.

**Clinical Utility:** Metabolomics plays an increasing role in clinical and translational research as large initiatives such as the Consortium of Metabolomics Studies (<https://epi.grants.cancer.gov/comets/>) and the NIH Common Funds Metabolomics Program (<https://commonfund.nih.gov/metabolomics>) are generating large-cohort metabolomics datasets (>1,000 participants). Because metabolomics profiles help define disease phenotypes and reflect alterations in the genome, epigenome, proteome, and environment (exposures and lifestyle), metabolites are ideal candidates for biomarker discovery in many diseases including cancer [37-39, 79-81]. With this in mind, metabolomics is playing a larger role in precision medicine, requiring continued efforts in data acquisition and analysis [82]. Metabolomics is also increasingly integrated with other

omics information and is analyzed in the context of biological pathways and networks, with the aim of identifying mechanisms that underlie diseases and finding novel therapeutic targets [34].

**Future Directions:** In October 2017, the NIH Common Fund has released funding opportunities to promote efforts in public accessibility and reuse of metabolomics data, development of computational tools for analysis (including omics integration) and interpretation of metabolomics data, and development of approaches to identify unknown metabolites. Thus, we anticipate further development in open-source, publicly available computational tools and infrastructures to facilitate metabolomics analysis. Since metabolomics is increasingly applied to biospecimens from large (>1000) cohorts and consortia, it is now possible to integrate other omics data, as well as clinical and environmental contexts in the analyses. The complexity of harmonizing data across cohorts and incorporating clinical and environmental data necessitates further standardization and computational infrastructure. Of special interest, the impact of alterations in the microbiome (dysbiosis) on metabolic pathways is particularly relevant since these dysbiosis-metabolome relationships can be causative or indicative of a myriad of human diseases [83, 84] including obesity and diabetes [85-87], cardiovascular diseases [88-90], inflammatory diseases [91], and cancer [64, 92, 93]. We thus suspect an increase in multi-tiered studies that apply a holistic approach to understanding diseases, including integration of molecular information from host and environment. Concurrently, as pathway information and identification of metabolites increases, strategies that take into account the kinetics of metabolites (e.g., metabolic flux and networks) will become more and more applicable for clinical metabolomics studies. Lastly, while the classical view of the molecular dogma is that metabolites levels are modulated by the epigenome, genome, and proteome, there have also been examples where metabolites regulate epigenetic events (i.e., going against the grain of the molecular dogma direction) [94-96]. The future of metabolomics and its potential in uncovering

biomarkers and deciphering mechanisms will surely necessitate modeling of complex bi-directional relationships within omics and environmental context information.

## Cell-free DNA

*Description of Technology:* Both normal and malignant cells shed DNA into the circulation and next-generation sequencing technologies are capable of detecting small amounts of cfDNA, making the blood a potential repository for tumor genomic profiling. „Liquid biopsy“, once validated, could enable the detection of cancer as a screening tool, track evidence for residual disease after cancer treatment, monitor patients for response to therapy, and discover meaningful mechanisms of resistance to cancer therapies. With this wealth of previously unavailable information, liquid biopsy could lead to the development of new assays, biomarkers, and targeted treatments to help cancer patients live longer, better lives. It is important to note that cell-free/circulating tumor DNA is only one aspect of ‘liquid biopsies,’ and there are multiple advances with other assays outside of the scope of this review, including circulating tumor cells [97-100], other nucleic acids [101], exosomes and other extracellular vesicles [102-104], and integrated biomarkers [105].

The presence of cell-free nucleic acids in the blood was first described in 1947 by Mandel and Métais [106] and three decades later, Leon et al. demonstrated that cancer patients had greater amounts of cfDNA relative to healthy controls [107]. Stroun et al. demonstrated both that tumor DNA was detectable specifically in plasma [108, 109] and that specific genomic alterations could be identified [110]. Of note, cfDNA is distinct and not derived from circulating tumor cells, although they are correlated and both increased in patients with advanced cancer [111]. Major advancements in cfDNA were first made in the field of perinatology, leading to the early minimally invasive detection of fetal chromosomal anomalies from maternal plasma in widespread clinical use today [112]. The remarkable advances in sequencing technology over the past two decades,

from Sanger sequencing to allele-specific PCR to the advent of massively parallel sequencing (‘next generation sequencing’) [113, 114], along with advances in bioinformatics analysis [115] and rapid reduction in cost have facilitated increasing ability to interrogate cfDNA to profile tumors.

*Recent Advancements:* To date, most clinical applications of cfDNA sequencing have focused on tracking specific mutations [111, 116-122] or sequencing targeted panels of cancer-related genes [123-127], particularly in metastatic cancer. In general, cfDNA is present in a greater proportion of patients and in larger amounts in metastatic cancers relative to primary tumors. In the metastatic setting, particularly in cancer types that are in many cases inaccessible (e.g., lung primary or metastases) or are higher risk lesions to sample in terms of potential complications, cfDNA genomic approaches may offer potential benefits relative to tumor biopsy. Tumors are known to be heterogeneous and biopsies inherently only sample a small localized region of a single metastatic site [128], introducing potential bias that may be overcome by cfDNA as a ‘sink’ of all metastatic sites in a patient [129]. Taking a patient-centered approach in the metastatic setting is critical – avoiding painful and inconvenient biopsies has the potential to improve quality of life. In one study, 34% of breast cancer patients undergoing metastatic biopsy described anxiety pre-biopsy and 59% described post-biopsy pain [130].

*Clinical Utility:* The only FDA-approved ‘liquid biopsy’ companion diagnostic to date is the cobas® EGFR Mutation Test v2 for the detection of exon 19 deletions or exon 21 (p.L858R) substitution mutations in the epidermal growth factor receptor (*EGFR*) gene to identify patients with metastatic non-small cell lung cancer eligible for treatment with erlotinib [131]. However, in cancers harboring mutations that are known to be prognostic or predictive, plasma-based cfDNA assays have demonstrated utility in disease management and are increasingly used clinically [132-134]. In addition, cfDNA targeted panel sequencing assays of cancer-related genes are used *in lieu* of met-

astatic tumor biopsy sequencing in clinical practice, including commercial tests such as Guardant360® and FoundationACT® [126, 135]. In the clinical setting, genomic profiling via cfDNA has been associated with more rapid turnaround of genomic results than tissue biopsies, frequently due to delays in accessing or obtaining tissue [136]. In the non-metastatic setting, there is great interest and excitement around the potential to develop patient tumor-specific panels of mutations for the highly sensitive detection of minimal residual disease after initial cancer treatment [137, 138]. In addition, multiple groups and commercial ventures are pursuing whether cfDNA could be used as a novel screening approach for cancer diagnosis [139], including the STRIVE Breast Cancer Screening Cohort (NCT03085888) supported by Grail, Inc. However, the optimal technical approach for cfDNA as a detection methodology remains unclear and large studies to assess sensitivity and specificity are only recently underway. Another approach is to incorporate cfDNA into a multi-analyte assay for cancer screening, such as CancerSEEK [105]. The CancerSEEK assay integrates a cfDNA PCR-based assay for a panel of common cancer mutations with established circulating protein biomarkers.

The promise of cfDNA is immense yet there remain several key unresolved challenges, including how well tumor-derived cfDNA mirrors tissue-derived tumor DNA, how to analyze tumor-normal DNA admixture present in circulation, how to better assess tumor-derived fraction of cfDNA, and how to account for clonal hematopoiesis of indeterminate potential (CHIP). While cfDNA appears to demonstrate overall high concordance with tumor biopsies [140-142], it is unclear whether cfDNA can serve as a comprehensive proxy for tumor biopsy in all contexts. Further, assays vary in their detection and reporting of genomic alterations from plasma [143].

*Data Challenges:* Circulating DNA in plasma is an admixture of both normal DNA shed primarily from leukocytes as well as tumor DNA, which presents challenges for analysis and interpretation of sequencing data. In the context of a large amount of tumor-derived DNA in the circulation (high

‘tumor fraction’), for example tumor fraction greater than 10%, standard next-generation sequencing approaches may be applied. However, in many contexts, tumor fraction is incredibly low particularly at diagnosis, in the setting of minimal residual disease, or in some ‘low cfDNA shedding’ cancer types and patient tumors. Highly specific assays may detect tumor fractions as low as 0.02% for panel sequencing and 0.00025% using such as digital droplet PCR for specific known alterations [138]. A major remaining challenge is to understand the sensitivity of assays for mutation detection to ensure that a negative test truly reflects the absence of tumor-derived DNA and not a limitation of assay or bioinformatic approaches.

As cfDNA assays seek to expand the breadth of sequencing (e.g., whole genome sequencing), efficient and cost-effective methods to screen blood samples for adequate amounts of tumor-derived DNA will be critical. Although sequencing costs continue to decline, identifying samples unlikely to provide usable sequence data should improve efficiencies. Most assays that determine tumor fraction depend on prior knowledge of tumor-specific mutations. Recent efforts suggest that low-coverage (approximately 0.1X) whole genome sequencing of cfDNA may offer the ability to quantify tumor fraction without the need for prior knowledge of tumor mutations [140].

Another challenge involves deconvolution of genomic alterations present in leukocytes as a consequence of CHIP from tumor-specific alterations [144, 145]. CHIP is the expansion of a clonal hematopoietic progenitor identified through common genomic alterations, present in increasing frequency as individual’s age. Typically, ‘normal’ DNA to distinguish germline from somatic tumor mutations is derived from peripheral blood cells and the frequency of CHIP – potentially more than 10% of patients over the age of 65 – suggests that methods to identify and account for CHIP will be critical.

Although most efforts to date have focused on tracking specific alterations known to be present in a tumor biopsy or sequencing targeted panels of cancer-related genes, there is growing evidence that cfDNA offers the potential to

obtain exome- and genome-level tumor sequencing data. Works from several groups have demonstrated the feasibility of genome-wide copy number analysis in cancer patients from plasma via shallow or low-coverage sequencing of cfDNA [140, 146-149]. Further efforts in this regard demonstrate feasibility of exome sequencing of cfDNA in the context of adequate tumor fraction [140-142, 146]. Comprehensive profiling is useful; particularly as blood can readily be collected serially, enabling tracking of the evolution of resistance as patients are on therapy. As we gain a greater understanding of the importance of non-driver mutations and regulatory elements in carcinogenesis and cancer progression, more comprehensive tumor genomic profiling from blood offers the potential for discovery in addition to detection, response tracking, or biomarker identification. In addition, more sensitive methods of detecting and isolating tumor-derived DNA or alterations from plasma may improve assay sensitivity [150, 151].

*Future Directions:* cfDNA is increasingly prevalent in oncology practice, from the first FDA-approved cfDNA biomarker to commercial cfDNA targeted panel sequencing assays. However, a recent American Society of Clinical Oncology (ASCO) and College of American Pathologists joint review reinforced that widespread use in clinical practice is not yet recommended until there is evidence of clinical validity and utility [152]. Despite this, there is growing evidence that personalized, highly sensitive mutation-based assays will be feasible for assessment of minimal residual disease and potentially tracking for early recurrence detection. These advances may translate to cfDNA assays that could be used for screening and early primary detection as well yet require clinical validation first. Finally, technological and computational advances are facilitating comprehensive genomic profiling exclusively from plasma. There remains the hope that new minimally invasive ‘liquid biopsy’ assays could improve outcomes by identifying cancer earlier and more specifically while also facilitating a greater understanding of novel susceptibilities and targets.

## Cohort Matching Algorithms

*Description of Technology:* Traditional biomarker analysis focuses on trying to figure out what distinguishes one patient from another patient. Broadly speaking, cohort-matching algorithms are either centered around similar features, or on similar outcomes. Using feature selection methods, biomarkers with the strongest association to the feature of interest are identified and then validated in an independent test set. These biomarker selection processes universally assume that there is a global, ground truth regarding the biomarker-phenotype relationship that is stable across multiple settings [153]. Unfortunately, this biomarker selection paradigm results in a tendency to divide patients into increasingly small subsets that may have no clinical relevance. Moreover, this fragmentation of previously “common” diseases results in a collection of “rare” subtypes that are then progressively challenging to study [154, 155] as there are an endless number of biomarker-subtype-therapy combinations. An alternative to this biomarker proliferation is the idea of trying to bin patients together based on potential outcome similarity – pattern-matching at a patient level. In other words, rather than focus on how patients are dissimilar, focus on how sets of patients respond similarly to a medication. In other words, one can leverage omics/phenomics comparisons at a patient level through more holistic pattern matching. This allows any number of omics technologies to define a patient-patient similarity strength.

*Recent Advances:* There currently is not a standard means of patient matching using omics data. There are an assortment of varied heuristics and cohort matching metrics [156-158]. Feature matching algorithms assume that retained features are critical determinants of outcomes such as survival and are optimal for situations where the biomarker is directly linked to the outcome. A straightforward approach to feature matching is to assign matches based on exact feature overlap -- for two patients to be a match they must share all features. Foundation Medicine’s PatientMatch tool [159] is an example of this exact matching approach. More complex feature matching schemes

have also been developed using Bayesian approaches [160]. Other feature matching algorithms include the PHeotypic Interpretation of Variants in Exomes (PHIVE) that matches human phenotypic profiles to glean the variants found in whole exome sequencing in mouse models [157] and DECIPHER [161] that enables international querying of karyotype, genetic, and phenotypic information for matches. In contrast to feature matching, the outcome-matching approach allows features to be weighted based on their discriminatory power. Frequently used algorithms are weighted K-nearest neighbor, random forest techniques, and deep learning (e.g., artificial neural networks) [162-164]. Outcomes matching attempts to match patients with other patients who may have a similar outcome to the same therapy based on phenotypic and omic predictors. A patient's features could potentially be compared not just from patient to patient (e.g., Patients Like Me) to infer outcomes but also from patient to cell lines (e.g., the Connectivity Map project [165]) and from a patient's electronic health record (EHR) to separate patient's EHRs [166].

*Clinical Utility:* The landscape today is dominated by feature matching strategies. These have been applied to clinical trial recruitment most notably for such national endeavors as the NCI-MATCH trial. Much of feature matching algorithms today have been focused on improving clinical trials accrual by prompting physicians to generate referrals [167]. Although seemingly simple, clinical trials matching algorithms have shown up to a 90% reduction in time spent identifying trial candidates [168]. GeneYenta matches phenotypically similar patients in regard to rare diseases [169] by weighting predictive features. Algorithms have been written to evaluate single nucleotide variant (SNV) frequencies between patients and non-small cell lung cancer cell lines to predict chemotherapeutic response [156]. Startup efforts such as MatchTX (<http://match-tx.com>) are attempting to reimagine social networking tools to help clinicians find best patient matches. Although the data sources, data types, and methods are heterogeneous, matching techniques at their core employ heuristic approaches to discover and vet the best profiles from large clinical databases.

*Data Challenges:* Cohort matching algorithms need to be capable of subsuming disparate data types and methods of comparison. Unfortunately the data types used in the matching process are varied and can be subjective or objective phenotypic measurements. Definitions of pathogenicity [170] remain a huge problem as do incomplete datasets and datasets lacking standardized ontologies. Preprocessing steps will need to be developed to organize the data into viable features to be used by matching algorithms. A further complication is the possibility that predictive models may require subsuming disparate unstandardized data-types simultaneously [63, 171]. EHR and omics interoperability remain a primary impediment to more robust algorithm generation. This will require a concerted standardization among data sets including vocabulary mapping and normalization.

*Future Directions:* As interoperability is a key impediment to the omics revolution, this has spurred efforts such as the Genomic Data Commons [172] which aims to “provide a unified data repository that enables data sharing across cancer genomic studies in support of precision medicine.” Other consortia efforts such as the Global Alliance for Genomics and Health (GA4GH) [173] and Health Level Seven International's Fast Healthcare Interoperability Resources (FHIR) [174] are enabling the development of application programming interfaces (APIs) and standards convergence. For example, the GA4GH Beacon Project allows federated queries to detect the existence of specific genomic variants across a variety of genome resources. Coalescing large datasets such that meaningful matching can occur has also been a thrust of recent developments. ASCO has also built a learning system called CancerLinQ [175] to help facilitate integration of data from multiple participating community oncology practice sites in an attempt to standardize data, facilitate research, and provide personalized cancer care through patient matching. Academic and selected larger oncology groups are participating in consortia such as ORIEN [173], GENIE [176], and the International Cancer Genome Consortium (ICGC) [177] and are building their respective frameworks for identifying patient cohorts. The “Sync for Science”

[178] endeavor sponsored by the NIH and the Office of the National Coordinator for Health Information Technology is going to permit patients to directly donate their data to be used to support innovative match-based algorithms for predictive purposes and thus contributing to precision medicine research. Sync for Science is also an integral part of the patient engagement portion of the NIH ‘All of Us’ initiative (<https://allofus.nih.gov>). Enhancing and perhaps complicating the field further, individual hospital systems such as the Swedish Cancer Institute and the Henry Ford Hospital system are developing their own precision medicine repositories. Commercial pathology laboratories – such as Caris and Foundation Medicine – have their internal datasets to mine. Other efforts like Syapse's Open Precision Oncology Network [179] allow aggregated cancer genomics data to be pulled from all participating health systems. These consortia and businesses all rely on patient matching as part of their core strengths.

## Conclusion

The sequencing of the genome has ushered in a new era of personalized cancer informatics. But the DNA genome is simply a first layer in a complex biological environment from which many omics data can be overlaid. We are in a time of growth. Metabolomics and proteomics are driving us closer to the tumor phenotype, and importantly, its response to treatment in real-time. ctDNA/cfDNA may help understand the clonal tumor evolution using non-invasive methods with the patient. These new omics datatypes will more than certainly help us tailor and adjust therapy for oncology patients. With these new datatypes and the understanding that data must be centralized, we are witnessing, too, an explosion of clinical/omics datasets aggregated by consortia and industry partners. As these datasets grow, so too, will be the need for more sophisticated cohort matching algorithms to bring clarity and useful actionable insights. These are exciting times. The cancer omics revolution continues to march forth rapidly and hopefully continues to improve our ability to practice precision oncology.

## References

- Goossens N, Nakagawa S, Sun X, Hoshida Y. Cancer biomarker discovery and validation. *Transl Cancer Res* 2015;4(3):256-69.
- U.S. Food and Drug Administration. Paving the way for personalized medicine: FDA's role in a new era of medical product development; 2013. Available from: [https://dx.advamed.org/sites/dx.advamed.org/files/resource/fda\\_report\\_on\\_paving\\_the\\_way\\_for\\_personalized\\_medicine.pdf](https://dx.advamed.org/sites/dx.advamed.org/files/resource/fda_report_on_paving_the_way_for_personalized_medicine.pdf). Accessed May 9, 2018.
- Catenacci DV. Next-generation clinical trials: Novel strategies to address the challenge of tumor molecular heterogeneity. *Mol Oncol* 2015;9(5):967-96.
- Akhmetov I, Bubnov RV. Assessing value of innovative molecular diagnostic tests in the concept of predictive, preventive, and personalized medicine. *EPMA J* 2015;6:19.
- Koren G. Personalized medicine-disregarding the obvious: analysis of trends among articles published on "Personalized Medicine". *Ther Drug Monit* 2015;37(5):559.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436-44.
- Brower V. NCI-MATCH pairs tumor mutations with matching drugs. *Nat Biotechnol* 2015;33(8):790-1.
- Bouquie R, Gregoire M, Hernando H, Azoulay C, Dailly E, Monteil-Ganiere C, et al. Evaluation of a methotrexate chemiluminescent microparticle immunoassay: comparison to fluorescence polarization immunoassay and liquid chromatography-tandem mass spectrometry. *Am J Clin Pathol* 2016;146(1):119-24.
- Yang JY, Yoshihara K, Tanaka K, Hatae M, Masuzaki H, Itamochi H, et al. Predicting time to ovarian carcinoma recurrence using protein markers. *J Clin Invest* 2013;123(9):3740-50.
- Cardnell RJ, Feng Y, Diao L, Fan YH, Masrorpour F, Wang J, et al. Proteomic markers of DNA repair and PI3K pathway activation predict response to the PARP inhibitor BMN 673 in small cell lung cancer. *Clin Cancer Res* 2013;19(22):6322-8.
- Sohn J, Do KA, Liu S, Chen H, Mills GB, Hortobagyi GN, et al. Functional proteomics characterization of residual triple-negative breast cancer after standard neoadjuvant chemotherapy. *Ann Oncol* 2013;24(10):2522-6.
- Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science* 2015;347(6220):12604-19.
- Li J, Akbani R, Zhao W, Lu Y, Weinstein JN, Mills GB, et al. Explore, visualize, and analyze functional cancer proteomic data using the Cancer Proteome Atlas. *Can Res* 2017;77(21):e51-e4.
- Vidova V, Spacil Z. A review on mass spectrometry-based quantitative proteomics: Targeted and data independent acquisition. *Anal Chim Acta* 2017;964:7-23.
- Hoofnagle AN, Becker JO, Oda MN, Cavigliolo G, Mayer P, Vaisar T. Multiple-reaction monitoring-mass spectrometric assays can accurately measure the relative protein abundance in complex mixtures. *Clin Chem* 2012;58(4):777-81.
- Mao X, He J, Li T, Lu Z, Sun J, Meng Y, et al. Application of imaging mass spectrometry for the molecular diagnosis of human breast tumors. *Sci Rep* 2016;6:21043.
- Vasaikar SV, Straub P, Wang J, Zhang B. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Res* 2017.
- Chen TW, Lee CC, Liu H, Wu CS, Pickering CR, Huang PJ, et al. APOBEC3A is an oral cancer prognostic biomarker in Taiwanese carriers of an APOBEC deletion polymorphism. *Nat Commun* 2017;8(1):465.
- Huang KL, Li S, Mertins P, Cao S, Gunawardena HP, Ruggles KV, et al. Proteogenomic integration reveals therapeutic targets in breast cancer xenografts. *Nat Commun* 2017;8:14864.
- Mertins P, Mani DR, Ruggles KV, Gillette MA, Clauser KR, Wang P, et al. Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* 2016;534(7605):55-62.
- Mundt F, Rajput S, Li S, Ruggles KV, Mooradian AD, Mertins P, et al. Mass spectrometry-based proteomics reveals potential roles of NEK9 and MAP2K4 in resistance to PI3K inhibitors in triple negative breast cancers. *Cancer Res* 2018.
- Zhang H, Liu T, Zhang Z, Payne SH, Zhang B, McDermott JE, et al. Integrated proteogenomic characterization of human high-grade serous ovarian cancer. *Cell* 2016;166(3):755-65.
- Tibes R, Qiu Y, Lu Y, Hennessy B, Andreeff M, Mills GB, et al. Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. *Mol Can Ther* 2006;5(10):2512-21.
- Desiere F, Deutsch EW, Nesvizhskii AI, Mallick P, King NL, Eng JK, et al. Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. *Gen Biol* 2005;6(1):R9.
- Farrar T, Deutsch EW, Hoopmann MR, Hallows JL, Sun Z, Huang CY, et al. The state of the human proteome in 2012 as viewed through PeptideAtlas. *J Proteome Res* 2013;12(1):162-71.
- Vizcaino JA, Cote RG, Csordas A, Dianes JA, Fabregat A, Foster JM, et al. The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nuc Acids Res* 2013;41(Database issue):D1063-9.
- Farrar T, Deutsch EW, Kreisberg R, Sun Z, Campbell DS, Mendoza L, et al. PASSEL: the PeptideAtlas SRMexperiment library. *Proteomics* 2012;12(8):1170-5.
- Lee JM, Hays JL, Annunziata CM, Noonan AM, Minasian L, Zujewski JA, et al. Phase I/II study of olaparib and carboplatin in BRCA1 or BRCA2 mutation-associated breast or ovarian cancer with biomarker analyses. *J Natl Cancer Inst* 2014;106(6):dju089.
- Endris V, Stenzinger A, Pfarr N, Penzel R, Mobs M, Lenze D, et al. NGS-based BRCA1/2 mutation testing of high-grade serous ovarian cancer tissue: results and conclusions of the first international round robin trial. *Virch Archiv* 2016;468(6):697-705.
- Matulonis UA, Penson RT, Domchek SM, Kaufman B, Shapira-Frommer R, Audeh MW, et al. Olaparib monotherapy in patients with advanced relapsed ovarian cancer and a germline BRCA1/2 mutation: a multistudy analysis of response rates and safety. *Ann Oncol* 2016;27(6):1013-9.
- Domchek SM, Aghajanian C, Shapira-Frommer R, Schmutzler RK, Audeh MW, Friedlander M, et al. Efficacy and safety of olaparib monotherapy in germline BRCA1/2 mutation carriers with advanced ovarian cancer and three or more lines of prior therapy. *Gyn Onc* 2016;140(2):199-203.
- Zhang A, Sun H, Xu H, Qiu S, Wang X. Cell metabolomics. *OMICS* 2013;17(10):495-501.
- LaConti JJ, Laiakis EC, Mays AD, Peran I, Kim SE, Shay JW, et al. Distinct serum metabolomics profiles associated with malignant progression in the KrasG12D mouse model of pancreatic ductal adenocarcinoma. *BMC Genomics* 2015;16 Suppl 1:S1.
- Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol* 2016;17(7):451-9.
- Hakimi AA, Reznik E, Lee CH, Creighton CJ, Brannon AR, Luna A, et al. An integrated metabolic atlas of clear cell renal cell carcinoma. *Cancer Cell* 2016;29(1):104-16.
- Clish CB. Metabolomics: an emerging but powerful tool for precision medicine. *Cold Spring Harb Mol Case Stud* 2015;1(1):a000588.
- Mathe EA, Patterson AD, Haznadar M, Manna SK, Krausz KW, Bowman ED, et al. Noninvasive urinary metabolomic profiling identifies diagnostic and prognostic markers in lung cancer. *Cancer Res* 2014;74(12):3259-70.
- Terunuma A, Putluri N, Mishra P, Mathe EA, Dorsy TH, Yi M, et al. MYC-driven accumulation of 2-hydroxyglutarate is associated with breast cancer prognosis. *J Clin Invest* 2014;124(1):398-412.
- Armitage EG, Southam AD. Monitoring cancer prognosis, diagnosis and treatment efficacy using metabolomics and lipidomics. *Metabolomics* 2016;12:146.
- Liesenfeld DB, Habermann N, Owen RW, Scalbert A, Ulrich CM. Review of mass spectrometry-based metabolomics in cancer research. *Cancer Epidemiol Biomarkers Prev* 2013;22(12):2182-201.
- Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. HMDB: the Human Metabolome Database. *Nucleic Acids Res* 2007;35(Database issue):D521-6.
- Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B, et al. HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res* 2009;37(Database issue):D603-10.
- Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res* 2013;41(Database issue):D801-7.
- Wishart DS. Emerging applications of metabolomics in drug discovery and precision medicine. *Nat Rev Drug Discov* 2016;15(7):473-84.
- Wishart DS. Advances in metabolite identification. *Bioanalysis* 2011;3(15):1769-82.
- Zhang A, Sun H, Wang P, Han Y, Wang X. Modern analytical techniques in metabolomics analysis. *Analyst* 2012;137(2):293-300.
- Dunn WB, Bailey NJ, Johnson HE. Measuring the metabolome: current analytical technologies. *Analyst* 2005;130(5):606-25.

48. Emwas AH. The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. *Methods Mol Biol* 2015;1277:161-93.
49. Roberts LD, Souza AL, Gerszten RE, Clish CB. Targeted metabolomics. *Curr Protoc Mol Biol* 2012;Chapter 30:Unit 30.2.1-24.
50. Vinayavekhin N, Saghatelian A. Untargeted metabolomics. *Curr Protoc Mol Biol* 2010;Chapter 30:Unit 30.1.1-24.
51. West JA, Beqqali A, Ament Z, Elliott P, Pinto YM, Arbustini E, et al. A targeted metabolomics assay for cardiac metabolism and demonstration using a mouse model of dilated cardiomyopathy. *Metabolomics* 2016;12:59.
52. Gelman SJ, Patti GJ. Profiling cancer metabolism at the 'omic' level: a last resort or the next frontier? *Cancer Metab* 2016;4:2.
53. Haznadar M, Mathe EA. Experimental and study design considerations for uncovering oncometabolites. *Methods Mol Biol* 2017;1513:37-47.
54. Dona AC, Kyriakides M, Scott F, Shephard EA, Varshavi D, Veselkov K, et al. A guide to the identification of metabolites in NMR-based metabolomics/metabolomics experiments. *Comput Struct Biotechnol J* 2016;14:135-53.
55. Watson DG. A rough guide to metabolite identification using high resolution liquid chromatography mass spectrometry in metabolomic profiling in metazoans. *Comput Struct Biotechnol J* 2013;4:e201301005.
56. Kind T, Fiehn O. Advances in structure elucidation of small molecules using mass spectrometry. *Bioanal Rev* 2010;2(1-4):23-60.
57. Grebe SK, Singh RJ. LC-MS/MS in the clinical laboratory - where to from here? *Clin Biochem Rev* 2011;32(1):5-31.
58. Li DW, Wang C, Bruschweiler R. Maximal clique method for the automated analysis of NMR TOCSY spectra of complex mixtures. *J Biomol NMR* 2017;68(3):195-202.
59. Bingol K, Bruschweiler R. Two elephants in the room: new hybrid nuclear magnetic resonance and mass spectrometry approaches for metabolomics. *Curr Opin Clin Nutr Metab Care* 2015;18(5):471-7.
60. Sud M, Fahy E, Cotter D, Azam K, Vadivelu I, Burant C, 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.
61. Salek RM, Haug K, Conesa P, Hastings J, Williams M, Mahendrakar T, et al. The MetaboLights repository: curation challenges in metabolomics. *Database (Oxford)* 2013;2013:bat029.
62. Menni C, Zierer J, Valdes AM, Spector TD. Mixing omics: combining genetics and metabolomics to study rheumatic diseases. *Nat Rev Rheumatol* 2017;13(3):174-81.
63. Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. Methods of integrating data to uncover genotype-phenotype interactions. *Nat Rev Genet* 2015;16(2):85-97.
64. Johnson CH, Spilker ME, Goetz L, Peterson SN, Siuzdak G. Metabolite and microbiome interplay in cancer immunotherapy. *Cancer Res* 2016;76(21):6146-52.
65. Alonso A, Marsal S, Julia A. Analytical methods in untargeted metabolomics: state of the art in 2015. *Front Bioeng Biotechnol* 2015;3:23.
66. Kelder T, Conklin BR, Evelo CT, Pico AR. Finding the right questions: exploratory pathway analysis to enhance biological discovery in large datasets. *PLoS Biol* 2010;8(8):e1000472.
67. Booth SC, Weljie AM, Turner RJ. Computational tools for the secondary analysis of metabolomics experiments. *Comput Struct Biotechnol J* 2013;4:e201301003.
68. Zhang B, Hu S, Baskin E, Patt A, Siddiqui JK, Mathe EA. RaMP: A comprehensive relational database of metabolomics pathways for pathway enrichment analysis of genes and metabolites. *Metabolites* 2018;8(1).
69. Siddiqui JK, Baskin E, Liu M, Cantemir-Stone CZ, Zhang B, Bonneville R, et al. IntLIM: integration using linear models of metabolomics and gene expression data. *BMC Bioinformatics* 2018;19(1):81.
70. Gowda H, Ivanisevic J, Johnson CH, Kurczy ME, Benton HP, Rinehart D, et al. Interactive XCMS Online: simplifying advanced metabolomic data processing and subsequent statistical analyses. *Anal Chem* 2014;86(14):6931-9.
71. Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G. XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal Chem* 2012;84(11):5035-9.
72. Xia J, Wishart DS. Using MetaboAnalyst 3.0 for comprehensive metabolomics data analysis. *Curr Protoc Bioinformatics* 2016;55:14.10.1-14.10.91.
73. Xia J, Sinel'nikov IV, Han B, Wishart DS. MetaboAnalyst 3.0--making metabolomics more meaningful. *Nucleic Acids Res* 2015;43(W1):W251-7.
74. Wanichthanarak K, Fan S, Grapov D, Barupal DK, Fiehn O. Metabox: A toolbox for metabolomic data analysis, interpretation and integrative exploration. *PLoS One* 2017;12(1):e0171046.
75. Vaniya A, Fiehn O. Using fragmentation trees and mass spectral trees for identifying unknown compounds in metabolomics. *Trends Analyt Chem* 2015;69:52-61.
76. Wohlgenuth G, Haldiya PK, Willighagen E, Kind T, Fiehn O. The Chemical Translation Service--a web-based tool to improve standardization of metabolomic reports. *Bioinformatics* 2010;26(20):2647-8.
77. Rocca-Serra P, Salek RM, Arita M, Correa E, Dayalan S, Gonzalez-Beltran A, et al. Data standards can boost metabolomics research, and if there is a will, there is a way. *Metabolomics* 2016;12:14.
78. Spicer R, Salek RM, Moreno P, Canueto D, Steinbeck C. Navigating freely-available software tools for metabolomics analysis. *Metabolomics* 2017;13(9):106.
79. Zhang A, Sun H, Yan G, Wang P, Wang X. Metabolomics for biomarker discovery: moving to the clinic. *Biomed Res Int* 2015;2015:354671.
80. Haznadar M, Cai Q, Krausz KW, Bowman ED, Margono E, Noro R, et al. Urinary metabolite risk biomarkers of lung cancer: a prospective cohort study. *Cancer Epidemiol Biomarkers Prev* 2016;25(6):978-86.
81. Patel S, Ahmed S. Emerging field of metabolomics: big promise for cancer biomarker identification and drug discovery. *J Pharm Biomed Anal* 2015;107:63-74.
82. Begler RD, Dunn W, Schmidt MA, Gross SS, Kirwan JA, Cascante M, et al. Metabolomics enables precision medicine: "A White Paper, Community Perspective". *Metabolomics* 2016;12(10):149.
83. Shaffer M, Armstrong AJS, Phelan VV, Reisdorff N, Lozupone CA. Microbiome and metabolome data integration provides insight into health and disease. *Transl Res* 2017;189:51-64.
84. Garg N, Luzzatto-Knaan T, Melnik AV, Caraballo-Rodriguez AM, Floros DJ, Petras D, et al. Natural products as mediators of disease. *Nat Prod Rep* 2017;34(2):194-219.
85. Devaraj S, Hemarajata P, Versalovic J. The human gut microbiome and body metabolism: implications for obesity and diabetes. *Clin Chem* 2013;59(4):617-28.
86. Boulange CL, Neves AL, Chilloux J, Nicholson JK, Dumas ME. Impact of the gut microbiota on inflammation, obesity, and metabolic disease. *Genome Med* 2016;8(1):42.
87. Sonnenburg JL, Backhed F. Diet-microbiota interactions as moderators of human metabolism. *Nature* 2016;535(7610):56-64.
88. Feng Q, Liu Z, Zhong S, Li R, Xia H, Jie Z, et al. Integrated metabolomics and metagenomics analysis of plasma and urine identified microbial metabolites associated with coronary heart disease. *Sci Rep* 2016;6:22525.
89. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 2011;472(7341):57-63.
90. Tang WH, Wang Z, Levison BS, Koeth RA, Britt EB, Fu X, et al. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med* 2013;368(17):1575-84.
91. Tong M, McHardy I, Ruegger P, Goudarzi M, Kashyap PC, Haritunians T, et al. Reprogramming of gut microbiome energy metabolism by the FUT2 Crohn's disease risk polymorphism. *ISME J* 2014;8(11):2193-206.
92. Contreras AV, Cocom-Chan B, Hernandez-Montes G, Portillo-Bobadilla T, Resendis-Antonio O. Host-microbiome interaction and cancer: potential application in precision medicine. *Front Physiol* 2016;7:606.
93. Sinha R, Ahn J, Sampson JN, Shi J, Yu G, Xiong X, et al. Fecal microbiota, fecal metabolome, and colorectal cancer interrelations. *PLoS One* 2016;11(3):e0152126.
94. Buescher JM, Driggers EM. Integration of omics: more than the sum of its parts. *Cancer Metab* 2016;4:4.
95. Lempradl A, Pospisilik JA, Penninger JM. Exploring the emerging complexity in transcriptional regulation of energy homeostasis. *Nat Rev Genet* 2015;16(11):665-81.
96. Etcheberry JP, Mostoslavsky R. Interplay between metabolism and epigenetics: a nuclear adaptation to environmental changes. *Mol Cell* 2016;62(5):695-711.
97. Alix-Panabieres C, Pantel K. Challenges in circulating tumour cell research. *Nat Rev Cancer* 2014;14(9):623-31.
98. Mohan S, Chemi F, Brady G. Challenges and unanswered questions for the next decade of circulating tumour cell research in lung cancer. *Transl Lung Cancer Res* 2017;6(4):454-72.
99. Polzer B, Medoro G, Pasch S, Fontana F, Zorzino

- L, Pestka A, et al. Molecular profiling of single circulating tumor cells with diagnostic intention. *EMBO Mol Med* 2014;6(11):1371-86.
100. Sparano JA, O'Neill A, Alpaugh K, Wolff AC, Northfelt DW, Dang C, et al. Abstract GS6-03: Circulating tumor cells five years after diagnosis are prognostic for late recurrence in operable stage II-III breast cancer. Presented at: San Antonio Breast Cancer Symposium. 2017 Dec 3-9; San Antonio, TX. (abstract)
  101. Kosaka N, Iguchi H, Ochiya T. Circulating microRNA in body fluid: a new potential biomarker for cancer diagnosis and prognosis. *Cancer Sci* 2010;101(10):2087-92.
  102. Garcia-Romero N, Esteban-Rubio S, Rackov G, Carrion-Navarro J, Belda-Iniesta C, Ayuso-Sacido A. Extracellular vesicles compartment in liquid biopsies: Clinical application. *Mol Aspects Med* 2018;60:27-37.
  103. Melo SA, Luecke LB, Kahler C, Fernandez AF, Gammon ST, Kaye J, et al. Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. *Nature* 2015;523(7559):177-82.
  104. Kalluri R. The biology and function of exosomes in cancer. *J Clin Invest* 2016;126(4):1208-15.
  105. Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* 2018;359(6378):926-30.
  106. Mandel P, Metais P. Les acides nucléiques du plasma sanguin chez l'homme. *C R Seances Soc Biol Fil* 1948;142(3-4):241-3.
  107. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ. Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res* 1977;37(3):646-50.
  108. Stroun M, Anker P, Lyautey J, Lederrey C, Maurice PA. Isolation and characterization of DNA from the plasma of cancer patients. *Eur J Cancer Clin Oncol* 1987;23(6):707-12.
  109. Stroun M, Anker P, Maurice P, Lyautey J, Lederrey C, Beljanski M. Neoplastic characteristics of the DNA found in the plasma of cancer patients. *Oncology* 1989;46(5):318-22.
  110. Vasioukhin V, Anker P, Maurice P, Lyautey J, Lederrey C, Stroun M. Point mutations of the N-ras gene in the blood plasma DNA of patients with myelodysplastic syndrome or acute myelogenous leukaemia. *Br J Haematol* 1994;86(4):774-9.
  111. Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin SF, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med* 2013;368(13):1199-209.
  112. Bianchi DW, Parker RL, Wentworth J, Madankumar R, Saffer C, Das AF, et al. DNA sequencing versus standard prenatal aneuploidy screening. *N Engl J Med* 2014;370(9):799-808.
  113. Garraway LA, Lander ES. Lessons from the cancer genome. *Cell* 2013;153(1):17-37.
  114. MacConaill LE. Existing and emerging technologies for tumor genomic profiling. *J Clin Oncol* 2013;31(15):1815-24.
  115. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* 2010;11(10):685-96.
  116. Silva JM, Silva J, Sanchez A, Garcia JM, Dominguez G, Provencio M, et al. Tumor DNA in plasma at diagnosis of breast cancer patients is a valuable predictor of disease-free survival. *Clin Cancer Res* 2002;8(12):3761-6.
  117. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. *Nat Med* 2008;14(9):985-90.
  118. Higgins MJ, Jelovac D, Barnathan E, Blair B, Slater S, Powers P, et al. Detection of tumor PIK3CA status in metastatic breast cancer using peripheral blood. *Clin Cancer Res* 2012;18(12):3462-9.
  119. Garcia-Murillas I, Schiavon G, Weigelt B, Ng C, Hrebien S, Cutts RJ, et al. Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Sci Transl Med* 2015;7(302):302ra133.
  120. Schiavon G, Hrebien S, Garcia-Murillas I, Cutts RJ, Pearson A, Tarazona N, et al. Analysis of ESR1 mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer. *Sci Transl Med* 2015;7(313):313ra182.
  121. Fribbens C, O'Leary B, Kilburn L, Hrebien S, Garcia-Murillas I, Beaney M, et al. Plasma ESR1 mutations and the treatment of estrogen receptor-positive advanced breast cancer. *J Clin Oncol* 2016;34(25):2961-8.
  122. Pawelz CP, Sacher AG, Raymond CK, Alden RS, O'Connell A, Mach SL, et al. Bias-corrected targeted next-generation sequencing for rapid, multiplexed detection of actionable alterations in cell-free DNA from advanced lung cancer patients. *Clin Cancer Res* 2016;22(4):915-22.
  123. Rothe F, Laes JF, Lambrechts D, Smeets D, Vincent D, Maetens M, et al. Plasma circulating tumor DNA as an alternative to metastatic biopsies for mutational analysis in breast cancer. *Ann Oncol* 2014;25(10):1959-65.
  124. Chae YK, Davis AA, Jain S, Santa-Maria C, Flaum L, Beaubier N, et al. Concordance of genomic alterations by next-generation sequencing (NGS) in tumor tissue versus circulating tumor DNA in breast cancer. *Mol Cancer Ther* 2017;16(7):1412-20.
  125. Forshe T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med* 2012;4(136):136ra68.
  126. Banks KC, Mortimer SAW, Zill OA, Lanman RB, Eltoukhy H, Talasz A. Abstract B140: Genomic profiling of over 5,000 consecutive cancer patients with a CLIA-certified cell-free DNA NGS test: Analytic and clinical validity and utility. *Mol Cancer Ther* 2015;14(12 Supplement 2):B140-B.
  127. Pearson A, Smyth E, Babina IS, Herrera-Abreu MT, Tarazona N, Peckitt C, et al. High-level clonal FGFR amplification and response to FGFR inhibition in a translational clinical trial. *Cancer Discov* 2016;6(8):838-51.
  128. Yates LR, Knappskog S, Wedge D, Farmery JHR, Gonzalez S, Martincorena I, et al. Genomic evolution of breast cancer metastasis and relapse. *Cancer Cell* 2017;32(2):169-84 e7.
  129. Siravegna G, Mussolin B, Buscarino M, Corti G, Cassingena A, Crisafulli G, et al. Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nat Med* 2015;21(7):795-801.
  130. Amir E, Miller N, Geddie W, Freedman O, Kassam F, Simmons C, et al. Prospective study evaluating the impact of tissue confirmation of metastatic disease in patients with breast cancer. *J Clin Oncol* 2012;30(6):587-92.
  131. cobas EGFR Mutation Test v2 2016. Available from: <https://www.fda.gov/Drugs/Information/OnDrugs/ApprovedDrugs/ucm504540.htm>. Accessed April 12, 2018.
  132. Sacher AG, Pawelz C, Dahlberg SE, Alden RS, O'Connell A, Feeney N, et al. Prospective validation of rapid plasma genotyping for the detection of EGFR and KRAS mutations in advanced lung cancer. *JAMA Oncol* 2016;2(8):1014-22.
  133. Thress KS, Pawelz CP, Felip E, Cho BC, Stetson D, Dougherty B, et al. Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. *Nat Med* 2015;21(6):560-2.
  134. Azad AA, Volik SV, Wyatt AW, Haegert A, Le Bihan S, Bell RH, et al. Androgen receptor gene aberrations in circulating cell-free DNA: Biomarkers of therapeutic resistance in castration-resistant prostate cancer. *Clin Cancer Res* 2015;21(10):2315-24.
  135. Zill OA, Greene C, Sebanovic D, Siew LM, Leng J, Vu M, et al. Cell-free DNA next-generation sequencing in pancreaticobiliary carcinomas. *Cancer Discov* 2015;5(10):1040-8.
  136. Pereira AAL, Morelli MP, Overman M, Kee B, Fogelman D, Vilar E, et al. Clinical utility of circulating cell-free DNA in advanced colorectal cancer. *PLoS One* 2017;12(8):e0183949.
  137. Tie J, Wang Y, Tomasetti C, Li L, Springer S, Kinde I, et al. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci Transl Med* 2016;8(346):346ra92.
  138. Newman AM, Bratman SV, To J, Wynne JF, Eclow NC, Modlin LA, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med* 2014;20(5):548-54.
  139. Phallen J, Sausen M, Adleff V, Leal A, Hruban C, White J, et al. Direct detection of early-stage cancers using circulating tumor DNA. *Sci Transl Med* 2017;9(403):eaan2415.
  140. Adalsteinsson VA, Ha G, Freeman SS, Choudhury AD, Stover DG, Parsons HA, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun* 2017;8(1):1324.
  141. Murtaza M, Dawson SJ, Pogrebniak K, Rueda OM, Provenzano E, Grant J, et al. Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nat Commun* 2015;6:8760.
  142. Murtaza M, Dawson SJ, Tsui DW, Gale D, Forshe T, Piskorz AM, et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* 2013;497(7447):108-12.
  143. Kuderer NM, Burton KA, Blau S, Rose AL, Parker S, Lyman GH, et al. Comparison of 2 commercially available next-generation

- sequencing platforms in oncology. *JAMA Oncol* 2017;3(7):996-8.
144. Genovese G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* 2014;371(26):2477-87.
  145. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med* 2014;371(26):2488-98.
  146. Heidary M, Auer M, Ulz P, Heitzer E, Petru E, Gasch C, et al. The dynamic range of circulating tumor DNA in metastatic breast cancer. *Breast Cancer Res* 2014;16(4):421.
  147. Ulz P, Auer M, Heitzer E. Detection of circulating tumor DNA in the blood of cancer patients: An important tool in cancer chemoprevention. *Methods Mol Biol* 2016;1379:45-68.
  148. Van Roy N, Van der Linden M, Menten B, Dheedene A, Vandeputte C, Van Dorpe J, et al. Shallow whole genome sequencing on circulating cell-free DNA allows reliable non-invasive copy number profiling in neuroblastoma patients. *Clin Cancer Res* 2017;23(20):6305-14.
  149. Stover DG, Parsons HA, Ha G, Freeman S, Barry WT, Guo H, et al. Genomewide copy number analysis of chemotherapy resistant metastatic triple-negative breast cancer from cell-free DNA. *J Clin Oncol* 2018;36(6):543-53.
  150. Ladas I, Fitarelli-Kiehl M, Song C, Adalsteinsson VA, Parsons HA, Lin NU, et al. Multiplexed elimination of wild-type DNA and high-resolution melting prior to targeted resequencing of liquid biopsies. *Clin Chem* 2017;63(10):1605-13.
  151. Sonnenberg A, Marciniak JY, Rassenti L, Ghia EM, Skowronski EA, Manouchehri S, et al. Rapid electrokinetic isolation of cancer-related circulating cell-free DNA directly from blood. *Clin Chem* 2014;60(3):500-9.
  152. Merker JD, Oxnard GR, Compton C, Diehn M, Hurley P, Lazar AJ, et al. Circulating tumor DNA analysis in patients with cancer: American Society of Clinical Oncology and College of American Pathologists joint review. *J Clin Oncol* 2018;Mar 5 [epub ahead of print].
  153. Lussier YA, Chen JL. The emergence of genome-based drug repositioning. *Sci Transl Med* 2011;3(96):96ps35.
  154. Johnson T LD, Chen JL. Opportunities for patient matching algorithms to improve patient care in oncology. *JCO Clinical Cancer Informatics* 2017;1(1).
  155. Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* 2014;158(4):929-44.
  156. Dudley JT, Chen R, Butte AJ. Matching cancer genomes to established cell lines for personalized oncology. *Pac Symp Biocomput* 2011:243-52.
  157. Robinson PN, Kohler S, Oellrich A, Sanger Mouse Genetics P, Wang K, Mungall CJ, et al. Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome Res* 2014;24(2):340-8.
  158. Wicks P, Vaughan TE, Massagli MP, Heywood J. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. *Nat Biotechnol* 2011;29(5):411-4.
  159. <http://investors.foundationmedicine.com/news-releases/news-release-details/foundation-medicine-launches-ice-2-its-new-version-interactive> (accessed May 8, 2018).
  160. Satagopan JM, Sen A, Zhou Q, Lan Q, Rothman N, Langseth H, et al. Bayes and empirical Bayes methods for reduced rank regression models in matched case-control studies. *Biometrics* 2016;72(2):584-95.
  161. Bragin E, Chatzimichali EA, Wright CF, Hurler ME, Firth HV, Bevan AP, et al. DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. *Nucleic Acids Res* 2014;42(Database issue):D993-D1000.
  162. Ma C, Ouyang J, Chen HL, Zhao XH. An efficient diagnosis system for Parkinson's disease using kernel-based extreme learning machine with subtractive clustering features weighting approach. *Comput Math Methods Med* 2014;2014:985789.
  163. Chen Y, Cao W, Gao X, Ong H, Ji T. Predicting postoperative complications of head and neck squamous cell carcinoma in elderly patients using random forest algorithm model. *BMC Med Inform Decis Mak* 2015;15:44.
  164. Belciug S, Gorunescu F. Error-correction learning for artificial neural networks using the Bayesian paradigm. Application to automated medical diagnosis. *J Biomed Inform* 2014;52:329-37.
  165. Lamb J. The Connectivity Map: a new tool for biomedical research. *Nat Rev Cancer* 2007;7(1):54-60.
  166. Mate S, Kopcke F, Toddenroth D, Martin M, Prokosch HU, Burkle T, et al. Ontology-based data integration between clinical and research systems. *PLoS One* 2015;10(1):e0116656.
  167. Embi PJ, Jain A, Clark J, Bizjack S, Hornung R, Harris CM. Effect of a clinical trial alert system on physician participation in trial recruitment. *Arch Intern Med* 2005;165(19):2272-7.
  168. Ni Y, Wright J, Perentesis J, Lingren T, Deleger L, Kaiser M, et al. Increasing the efficiency of trial-patient matching: automated clinical trial eligibility pre-screening for pediatric oncology patients. *BMC Med Inform Decis Mak* 2015;15:28.
  169. Gottlieb MM, Arenillas DJ, Maithripala S, Maurer ZD, Tarailo Graovac M, Armstrong L, et al. GeneYenta: a phenotype-based rare disease case matching tool based on online dating algorithms for the acceleration of exome interpretation. *Hum Mutat* 2015;36(4):432-8.
  170. Hansen NF. Variant calling from next generation sequence data. *Methods Mol Biol* 2016;1418:209-24.
  171. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods* 2014;11(3):333-7.
  172. Jensen MA, Ferretti V, Grossman RL, Staudt LM. The NCI Genomic Data Commons as an engine for precision medicine. *Blood* 2017;130(4):453-9.
  173. Cook-Deegan R, Ankeny RA, Maxson Jones K. Sharing data to build a medical information commons: from Bermuda to the Global Alliance. *Annu Rev Genomics Hum Genet* 2017;18:389-415.
  174. Alterovitz G, Warner J, Zhang P, Chen Y, Ullman-Cullere M, Kreda D, et al. SMART on FHIR Genomics: facilitating standardized clinico-genomic apps. *J Am Med Inform Assoc* 2015;22(6):1173-8.
  175. Miller RS. CancerLinQ Update. *J Oncol Pract* 2016;12(10):835-7.
  176. Consortium APG. AACR Project GENIE: Powering precision medicine through an international consortium. *Cancer Discov* 2017;7(8):818-31.
  177. Joly Y, Dove ES, Knoppers BM, Bobrow M, Chalmers D. Data sharing in the post-genomic world: the experience of the International Cancer Genome Consortium (ICGC) Data Access Compliance Office (DACO). *PLoS Comput Biol* 2012;8(7):e1002549.
  178. Turvey C, Klein D, Fix G, Hogan TP, Woods S, Simon SR, et al. Blue Button use by patients to access and share health record information using the Department of Veterans Affairs' online patient portal. *J Am Med Inform Assoc* 2014;21(4):657-63.
  179. <https://www.syapse.com/news/press-releases/oncology-precision-network-open-announces-data-sharing-commitments-at-vice-president-bidens-cancer-moonshot-summit> (accessed May 8, 2018)

Correspondence to:  
James L Chen, MD  
Division of Medical Oncology and  
Department of Biomedical Informatics  
The Ohio State University  
320 W 10<sup>th</sup> Ave  
Columbus, OH, 43210  
Tel: +1 614 685 9754  
E-mail: James.Chen@osumc.edu