MULTI-OMICS APPROACH FOR BIOMARKER DISCOVERY IN CANCER

Abstract

Despite significant progress in the recent past, cancer remains as an enigma for both clinicians and scientists alike and the underlying mechanism(s) of onset and progression of the disease remain unresolved. Nevertheless, in recent years several efforts have been crafted for use of high through put omics technologies including genomics (detection of cancer- specific mutations across the genome), epigenomics (identification of modified epigenetic regulations within cancerous cells), transcriptomics (scrutinization of differential expression of mRNAs), proteomics (exploration of differentially expressed proteins) and metabolomics (comprehensive analysis of small metabolites) for unraveling reliable biomarkers for cancer. Isolated omics subset data may offer an incomplete, skewed and biased picture while integrating different omics techniques is likely to evaluate the cross-talk between various aspects of the complicated physiology of cancer initiation and development. Integration of multi-omics data from different molecular domains with clinical data and epidemiologic risk factors may provide a sensitive, accurate and reproducible technique for interpreting the intricacies of cancer. Hence, a detailed analysis of multiomics data in an integrated framework is hypothesized to accelerate identification of potential cancer biomarkers leading to timely diagnosis and prompt medical intervention in the field of precision medicine and is the major focus of this chapter.

Keywords: OMICS; Cancer; Biomarker; Data Integration

Authors

Amrita Swain
Department of Zoology
Biochemistry and Molecular Biology Laboratory School of Life Sciences
Centre for Excellence in Environment and Public Health
Ravenshaw University
Odisha, India.

Anwesha Pradhan
Department of Zoology
Biochemistry and Molecular Biology Laboratory School of Life Sciences
Centre for Excellence in Environment and Public Health
Ravenshaw University
Odisha, India.

Soumya Ranjan Jena
Department of Zoology
Biochemistry and Molecular Biology Laboratory School of Life Sciences
Centre for Excellence in Environment and Public Health
Ravenshaw University
Odisha, India.

Dhananjay Soren
Department of Zoology
Biochemistry and Molecular Biology Laboratory School of Life Sciences
Centre for Excellence in Environment and Public Health
Ravenshaw University
Odisha, India.

Luna Samanta
Department of Zoology
Biochemistry and Molecular Biology Laboratory School of Life Sciences
Centre for Excellence in Environment and Public Health
Ravenshaw University
Odisha, India.
lsamanta@ravenshawuniversity.ac.in

I. INTRODUCTION

Cancer, a highly lethal and complex chronic disease, poses a significant global challenge, causing millions of deaths annually and impacting life expectancy negatively. Recent reports from GLOB-CAN indicate a rising incidence of 19.3 million cases and 10 million cancer-related deaths worldwide, showing a steady increase from 2012 to 2020 (Sung et al. 2021). Despite significant advances in diagnostics and therapeutics, cancer remains a perplexing challenge for both biologists and clinicians. While past decades held the dream of conquering this disease, current research in cancer management has achieved considerable success. However, many of these breakthroughs are tailored to individual patients. For example, evidence-based guidelines recommend colorectal cancer screening for individuals aged 50 and older, but those with Lynch syndrome (affecting 1 in 280 individuals) face an elevated risk at a younger age, necessitating a more targeted approach (Yurgelun and Hampel 2018). To fully overcome cancer, new therapeutic approaches with enhanced mechanisms are required. Researchers must not only focus on improved treatments but also on methods for predicting an individual's cancer risk, detecting cancer at an early stage, distinguishing aggressive from non-aggressive cancers, and monitoring recurrence and treatment response. Cancer can originate through various pathways in different cells at varying rates. Identifying the clinical, molecular, and genetic events in these pathways can lead to the development of preventive strategies, including the creation of biomarkers that may be observable before clinical cancer detection.

II. CLINICAL BIOMARKERS

A biomarker is a distinct molecular signature that indicates a physiological condition, allowing for objective measurement to distinguish between normal and pathological states or to assess the response to therapy. Biomarkers have a rich history in clinical practice, ranging from simple pulse rate measurements, blood tests, and X-rays to more intricate laboratory assessments. They encompass various types, including genomic (DNA and RNA), protein, metabolite, carbohydrate, imaging, and cellular biomarkers, often linked to specific pathological elements, histological or radiographic properties, and genetic alterations. An effective biomarker is tailored for a particular disease state and can be readily quantified using bodily fluids like serum, saliva, urine, or cerebrospinal fluid. Molecular markers have evolved as valuable tools for disease diagnosis, epidemiological studies, and health-related assessments, spanning diverse fields from cancer to inflammatory, neurological, and cardiovascular diseases, owing decades of research . For example, HbA1c serves as a widely used biomarker for prediabetes and diabetes, while C-reactive protein (CRP) is employed as a biomarker for inflammation (Dorcely et al. 2017).

1. **Established biomarkers of Cancer:** Genomic profiling technologies and selective molecular targeted therapies represent recent advancements in cancer management, underscoring the pivotal role of biomarkers (Goossens et al., 2015). Cancer biomarkers, including proteins, DNA, RNA, and metabolites, have diverse roles in oncology, such as screening, monitoring, risk assessment, diagnosis, recurrence prediction, and prognosis. They are crucial for customizing treatment plans and managing drug reactions. These biomarkers can also classify cell types and aid in dose- response studies. They are categorized into three types: predictive, prognostic, and diagnostic, based on their application approach (Conley and Taube, 2004). Predictive biomarkers offer insights into a patient's likely response to specific therapies, such as the increased expression of the

HER2/neu protein predicting trastuzumab response in breast cancer. Prognostic biomarkers, on the other hand, provide warnings about potential long-term outcomes, such as cancer recurrence or disease progression. Examples include elevated Prostate-specific antigen (PSA) levels in prostate cancer, Chromosome 17p deletions, and TP53 mutations in Chronic myeloid leukemia (CML) patients, as well as BRCA1 and BRCA2 gene mutations. Diagnostic biomarkers, including the Bence–Jones protein urine test for multiple myeloma, carcinoembryonic antigen levels in colorectal cancer (CRC) surveillance, PSA level measurement in prostate cancer, and CDC 20 usage in diagnosing and treating relapsed and refractory lymphoma, aid in disease identification. Cancer often arises from genetic or epigenetic changes that lead to alterations in protein expression due to post-translational modifications. These changes impact cell progression, apoptosis, and the secretion of factors affecting neighboring cells (Maruvada et al., 2005). Molecular biomarkers, accessed through genes, genetic alterations, mRNA, and protein differential expression, encompass a wide range of biochemical entities, including proteins, nucleic acids, sugars, small metabolites, and more (Maruvada et al., 2005). As a result, biomarkers are categorized into genetic, transcriptomic, epigenetic, proteomic, and metabolomic types based on their level of molecular alterations. Genetic biomarkers, detectable through liquid biopsies and blood samples, aid in prognosis and cancer identification. Examples include BRAFV600VE mutations in melanoma and ALK gene rearrangements in lung cancer. Transcriptomic markers, like KAT2B, PCNA, CD86, miR-192-5p, and miR-215-5p in cervical cancer, as well as RNY3P1, RNY4P1, and RNY4P25 overexpression in melanoma patients, provide insights into gene expression patterns (Kori and Yalcin Arga, 2018). Epigenetic biomarkers, such as APC, GSTP1, and RARβ2 promoter methylation for detecting prostate cancer in urine, as well as SHOX2 and CDKN2A promoter methylation in lung cancer, elucidate epigenetic modifications (Kori and Yalcin Arga, 2018). Proteins serve as major biomarkers due to their crucial roles in cellular function and metabolism. Abnormally expressed proteins often underlie diseases, especially cancer. Detecting protein- based markers requires stringent and specific techniques, including ELISA, electrochemical, electrical, and optical methods (Wu and Qu, 2015). Prominent examples of protein biomarkers include EpCAM, CD45, and cytokeratins 8, 18, and 19, which aid in detecting circulating tumor cells (CTCs) and monitoring patients (Yousefi et al., 2021). Additionally, proteins like HE4 and CA-125 are measured to assess the risk of ovarian malignancy using algorithms (Fujiwara et al., 2015). Metabolomic biomarkers, unique to metabolic pathways, vary in different cancer types. For instance, breast cancer patients exhibit decreased lysophosphatidyl ethanolamine (LPE) levels and increased ceramide levels, while lung cancer patients show reduced choline and linoleic acid in their serum. Gastric cancer patients may have elevated 3-hydroxypropionic acid levels and reduced pyruvic acid levels in their serum. Nucleic acid biomarkers involve the measurement of DNA and RNA, providing crucial insights into cancer biology and progression. While these biomarkers have been instrumental in cancer management for years, cancer remains a complex challenge with diverse aberrant networks that may offer new potential targets. Beyond aiding clinical decision-making, cancer biomarkers are closely linked to deregulated molecular pathways and cancer etiology, thus validating specific treatment options. The current revolution in high-throughput gene sequencing and increased molecular characterization is driving the analysis of complex cancer mechanisms in a personalized manner, leading to the concept of precision medicine (Goossens et al., 2015). This approach recognizes individual variability, encompassing alterations in DNA, RNA, proteins, and metabolites, as integral components of prevention and treatment strategies.

2. **Precision medicine and omics:** In cancer, the tremendous variability among patients underscores the need for personalized approaches. Traditional biomarkers may lose diagnostic reproducibility over time, necessitating robust validation since it needs availability of specimens, assay platforms and study design (Goossens et al. 2015). Precision medicine tailors disease prevention, diagnosis, and treatment to each patient's unique characteristics taking into account differences in people's genes, environments, and lifestyles. Omics technologies have become integral to precision medicine because they allow for a deep characterization of an individual's biological makeup. By analyzing a person's genome (genomics), gene expression (transcriptomics), protein profiles (proteomics), and metabolic pathways (metabolomics), omics generates diverse data sets, making it a promising tool for discovering precise cancer biomarkers.

3. **Omics approach to biomarker:** Cancer biomarker research advances alongside technology, with omics approaches proving to be powerful tools for understanding the complex cellular dysfunctions driving cancer development (Alyass et al., 2015). Molecular changes in cancer occur across various levels, such as the genome, proteome, epigenome, metabolome, and transcriptome, which are crucial for understanding cancer characteristics. Omics tools, like proteomics, genomics, metabolomics, transcriptomics, and radiomics, are employed to comprehensively analyze these elements. Recent research has focused on using single omics methods to uncover molecular mechanisms in cancer development, involving genome scanning, epigenetic studies, and investigating mRNA and protein expression changes. Next-generation sequencing (NGS) has significantly improved genomic research efficiency, leading to increased identification of genetic factors in cancer. These approaches provide valuable insights into cancer genomes and genes related to tumorigenesis (Roberts et al., 2012). Genome profiling has the ability to recognize various molecular subtypes and categorize patients, a vital aspect of personalized medicine. To discover new biomarkers in cancer proteomics and metabolomics, techniques like high-performance liquid chromatography (HPLC), nuclear magnetic resonance (NMR), and mass spectrometry (MS) are frequently used (Alessandro et al., 2005). These biomarkers serve various purposes, including prognostic biomarkers for estimating patient clinical outcomes, predictive biomarkers for categorizing treatment options, and diagnostic biomarkers for early detection, all of which are vital for cancer forecasting and prevention. A notable shift is occurring in cancer treatment, characterized by predictive, preventive, and personalized medicine (PPPM), intricately tailored through single omics approaches.

4. **The omics platforms:** Omics is a contemporary analytical approach widely applied to biological samples, encompassing diverse fields like genomics, proteomics, metabolomics, transcriptomics, and more, generating extensive datasets focused on specific biomolecules. Genomics, transcriptomics, proteomics, and metabolomics, in particular, show promise for pharmaceutical applications, especially in cancer therapy. However, the challenge lies in translating complex omics data into meaningful biological insights. Various terms such as multi-omics, trans-omics, and integrated omics are used to describe these approaches, each with distinct data processing and interpretation methods. Omics methodologies heavily rely on integration and interpretation to advance our understanding of systems biology. Emerging interdisciplinary analyses like glycomics and microbiomics aim to streamline processes and reduce costs, facilitating comprehensive studies across various biological and computational disciplines. Transcriptomics, proteomics, and metabolomics are commonly employed in the study of

biological networks. Successful integration of omics data can yield additional insights into complex biological networks, including information on allosteric regulation, protein-protein interactions, and more. Omics data can be linked to phenotypic characteristics to uncover genetic and environmental factors. Case studies involving rats and bacteria have demonstrated positive correlations between different omics layers, leading to the discovery of pathway-wide phenome-wide associations and trans-ome-wide associations (Trans-OWAS) (Misra et al. 2019). Repositories like The Cancer Genome Atlas (TCGA) offer open access to omic data from various cancer types, facilitating data retrieval for researchers (Fernandez- Lozano et al. 2018). The challenge remains in translating multi-dimensional omics data into biologically relevant contexts. In the realm of omics, genomics, epigenomics, transcriptomics, proteomics, and metabolomics are pivotal for advancing pharmaceutical research, particularly in cancer therapy (Figure 1).

5. **Genomics:** Genomic techniques are primarily deployed to sequence an individual's genome in order to better understand interindividual differences at both the somatic and germ line levels through (SNPs) single nucleotide polymorphisms, loss of heterozygosity variants, copy number variants (CNVs), genomic rearrangements, and rare variants), insertions and deletions (INDELs). Researchers have been able to sequence the genome/exome of choice extensively enough define the mutational patterns of a given tissue owing to the ultimate progression of sequencing technology from Sanger sequencing-based techniques to Next Generation Sequence (NGS)-based massively parallel sequencing (Chakraborty et al. 2018). High-speed and high-throughput NGS technologies greatly enhances cancer genome analysis and demonstrates the entire spectrum of cancer gene mutation, which not only is used to lead the discovery of novel targeted drugs, but also has a huge effect on understanding cancer biology and accelerating approaches for PPPM in cancer (Lu and Zhan 2018). Genomics is implemented in multiple areas including biotechnology, developmental biology, diagnostics and therapeutics, pharmaceutical industry, gene therapy, disease prevention, comparative genomics and evolutionary biology.

6. **Epigenomics:** Epigenomics involves the systematic examination of reversible epigenetic modifications, a widely adopted approach worldwide. It encompasses alterations in gene expression resulting from modifications to a cell's DNA or histone proteins without changing the underlying DNA sequence (Wang and Chang 2018). Notable epigenomic signatures include DNA methylation, acetylation, histone modification, and the precise positioning of nucleosomes. These signatures can serve as potential biomarkers for various diseases, including cancer. Gene expression regulation, particularly the modification of expression 88^{th} gene, is a prominent focus in modern omics research across diverse fields. Modified epigenomic alterations can lead to clinical phenotypic variations. The complexity of cancer, characterized by extensive variations in the expression of multiple genes, is further compounded by factors such as germline genetic factors, somatic mutations, and epigenetic contributors that can influence these changes. Single-cell gene expression analysis allows for the assessment of cell-specific somatic mutations on gene expression in individual cells. Epigenomics employs techniques like Chromatin Immunoprecipitation Sequencing (ChIP-seq) and DNA methylation analysis through whole-genome bisulfite or array-based sequencing. ChIP-seq, a crucial epigenomic tool, is utilized to identify DNA-binding sites for transcription factors (TFs) and histone proteins, potentially validating them as biomarkers with comprehensive genome-wide profiles. DNA methylation analysis involves digestion assays and bisulfite

sequencing of DNA, where methylation-sensitive and -insensitive restriction enzymes are used for genomic DNA digestion. Cytosines are converted to uracil during bisulfite treatment, preserving the methylated cytosine. The generated DNA is then analyzed through techniques such as PCR, MALDI-TOF mass spectrometry, or microarrays. To further advance these analyses, various bioinformatic tools, such as Biopearl, Biojava, Biopython, and machine learning approaches, are employed (Bayón et al. 2016). Whole-genome analyses, known as Genome-Wide Association Studies (GWAS), aim to identify the impact of Single Nucleotide Variants (SNVs) on phenotypic traits, while Epigenome-Wide Association Studies (EpWAS) are designed to investigate the influence of epigenomic variations.

7. **Transcriptomics:** Transcriptome encompasses all RNA molecules, including mRNA, rRNA, tRNA, and non-coding RNAs, produced in cells, reflecting genetic and environmental influences (Yan et al. 2015). Transcriptomics profiles an organism's complete RNA set, which plays diverse roles in cellular functions. Two primary transcriptomics techniques are microarrays and RNA sequencing (RNA-seq), with RNA-seq gaining prominence due to its independence from prior organism information and lower sample requirements. It involves RNA-based omics, comprising steps such as sample collection, RNA extraction, clonal amplification, library preparation, and sequencing (e.g., pyrosequencing). Subsequent workflow involves cleaning, screening, alignment (reference-based or de novo), variant calling, annotation, functional prediction, and pathway analyses (Misra et al. 2019). RNA-seq enhances sensitivity by targeting specific RNA, requiring fragmentation, conversion to complementary DNA (cDNA) using reverse transcriptase and PCR amplification. Sequencing can be single-end or paired-end, with the latter facilitating more assemblies and alignments for gene annotation. Nanopore sequencing is an emerging method that bypasses amplification but is still developing. Bioinformatics tools like FaQCs and FastQC analyze raw RNA-seq data for sequence quality, contamination, k-mer identification, and errors. Alignment software aligns transcript sequences to a reference genome, followed by quantification at the gene, exon, or transcript level. Data normalization and statistical analysis, using software like EdgeR, Cuffdiff2, Limma/Voom, or DEseq2, reveal differential gene expression. Validation is often conducted through qPCR for target and control gene expression (Conesa et al. 2016).

8. **Proteomics:** Transcriptomics platforms are limited in their ability to reveal the functional aspects of gene expression, as not all RNAs translate into proteins, and some undergo post-translational modifications. To address this, proteomics has emerged, focusing on identifying and quantifying the entire protein content to find disease biomarkers. Traditional proteomics techniques include chromatography and western blotting (Dalal et al. 2020). Recent advancements have introduced shotgun and targeted approaches, allowing for the quantification of collective protein samples. Mass Spectrometry (MS) has played a pivotal role in enhancing sensitivity and data analysis in proteomics, even with smaller sample sizes. This technique can detect subtle differences in amino acid sequences, protein abundances, post-translational modifications (PTMs), PTM sites, and more. Proteomics workflows involve sample collection, protein extraction, enzymatic digestion into peptides, liquid chromatography (LC) fractionation, followed by Mass Spectrometry for protein and peptide identification and quantification. Bioinformatics analyses, such as network and pathway analyses, are also conducted. LC methods, particularly 2DLC and multi-dimensional LC (MDLC), are commonly used for proteomic

fractionation/separation. Stable isotopes (e.g., iTRAQ and TMT labeling) paired with 2DLC enable proteome component characterization (Geng et al. 2009). Various MS types, including MALDI-TOF and Fourier transform ion cyclotron resonance (FTICR), are effective for identifying protein expressions, PTMs, variations, and species (Mao et al. 2013). The data analysis process includes converting raw LC-MS/MS files to mzML peak lists, followed by analysis using databases such as Myrimatch, Pepitome, and MS-GF+ (Ma et al., 2011). IDP3 is then used to convert identified spectral files into IDPicker 3 SQLite Databases (idpDB) files and gather these idpDB files. Additionally, Swissprot filters MS data raw files against the human proteome database using the Sequest HT algorithm. Classification systems like PANTHER, GO annotation, and ingenuity pathway analysis (IPA) are employed for data mining, functional analysis, and pathway analysis of identified proteins (Dalal et al. 2020).

9. **Metabolomics:** Metabolomics combines various omics fields like genomics, transcriptomics, and proteomics to assess an organism's phenotype. It focuses on small molecular weight metabolites, the end products of complex biochemical processes, ranging from 50 to 1500 Dalton. The metabolome encompasses the total collection of metabolites in a biological sample related to metabolism (Dalal et al. 2020). Metabolomic analyses involve quantifying small metabolites like lipids, sugars, amino acids, nucleic acids, drugs, and steroids in various sample types, including cell lines, tissues, biofluids, and different environmental conditions. Techniques such as NMR spectroscopy and chromatography coupled to mass spectrometry (LC-MS and GC-MS) are employed based on the specific application and instrumentation requirements. Data analysis includes preprocessing, statistical analysis, pattern recognition, and the use of databases like METLIN, HMDB, and KEGG. Metabolomics analysis stages consist of experiment design, sample collection, metabolism quenching, metabolite extraction, optional chemical derivatization, analytical techniques (MS or NMR), and data analysis involving filtering, alignment, statistical analysis, imputation, annotation, and network/pathway analysis. Each step may vary depending on the research objectives, data type, and instrumentation used (Misra et al. 2019). While single omics approaches provide valuable insights, they are insufficient for understanding complex diseases like cancer, which involve intricate biological interactions. Incorporating multi-omics data is crucial for unraveling the intricate molecular signatures of tumorigenic networks and discovering new biomarkers and drug targets.

III. MULTI-OMIC APPROACH AND DATA INTEGRATION

Extracting meaningful correlations and real interactions from vast omics datasets is a computationally challenging task due to non-linear interactions and collective effects in biological systems. Differentiating genuine biological signals from random noise and irrelevant analytical systems poses difficulties. High-dimensional datasets exhibit variations in gene, protein, and metabolite expression across individuals, organs, tissues, and cells, further complicating data extraction. There are various methods for integrating multidimensional omics data, often focusing on mutations to identify genomic determinants of phenotypic features and distinguish driver from passenger mutations (Yu and Zeng 2018). Two integration approaches exist: bottom-up and top-down. The bottom-up approach consolidates various data types first and then manually integrates distinct clusters, while top-down techniques combine all data types simultaneously, allowing for data integration and dimensionality reduction together. Data integration methods encompass regression,

exploratory, predictive, unsupervised, semi-supervised, or supervised analyses. Unsupervised models derive deductions from input factors without a marked response variable (Huang et al. 2017). Various algorithms, such as multivariate, Bayesian network-based, fusion-based, correlation-based, and similarity-based methods, can be employed for data integration. Figure 2 illustrates several tools used for multi-omics data integration.

1. **Multivariate methods:** Unsupervised multi-omics data integration often employs joint non- negative matrix factorization (NMF), which splits non-negative matrices into two component matrices without negative elements. This technique projects various data types onto a shared coordinate frame, facilitating examination. For instance, integrating microRNA, mRNA expression, and DNA methylation data in TCGA ovarian cancer samples with NMF revealed new signaling pathway alterations and patient subtypes (Zhang et al. 2012). Another method, JIVE (Joint and Individual Variation Explained), decomposes variations in datasets into three terms: joint variation shared across datasets, specific structured variation within datasets, and residual noise. While JIVE provides more precise evaluation of common characteristics, it can be affected by outliers. In glioblastomas, JIVE combining miRNA and gene expression data improved tumor type identification (Lock et al. 2013). MoCluster employs multivariate analysis to detect similar patterns across diverse omics datasets, followed by a clustering algorithm to identify distinct clusters. This approach identified microsatellite instability-high tumors and three novel subgroups of colorectal cancer by incorporating mRNA, protein, and methylation data (Meng et al. 2016).

2. **Statistical methods:** The Bayesian algorithm accommodates different datasets with varying dispersions and associations. iCluster, a substantial clustering technique, employs a Gaussian latent variable model to integrate multiple genetic traits (Shen et al., 2009). It focuses on obtaining diverse sample clusters and identifying associated characteristics. Integrating copy number variants and gene expression profiles using unsupervised clustering identified novel breast cancer molecular subtypes with unique therapeutic implications (Curtis et al. 2012). iClusterPlus, an improved version, performs model-based matrix factorization integration. It decomposes each omics data type into component and loading factors, revealing gene characteristics and latent cancer subgroups. iCluster+ is effective for tumor classification and biomarker discovery in cancer genomics, such as MYB and PCM1 in leukemia and Scotin, BAP1, and XPC in small-cell lung cancer (Mo et al. 2013). However, it requires parameter tuning, substantial computations, and lacks statistical significance assessment for selected variables (Mo et al. 2018). The Bayesian Consensus Clustering (BCC) approach focuses on Finite Dirichlet mixture models to explore clusters within individual datasets and integrate them (Lock and Dunson 2013). Integrative Bayesian analysis of genomics data (iBAG), a supervised multiblock technique, assesses clinical associations between omics data from various platforms and identifies biomarkers linked to clinical outcomes. It combines clinical data, survival statistics, and omics components to uncover methylation-regulated genes related to patient survival in glioblastoma samples (Wang et al. 2013).

3. **Network-based integration:** iOmicsPASS enables supervised integration of DNA copy number, mRNA, and protein expression data to construct interconnected subnetworks using a modified closest shrunken centroid technique for accurate phenotypic group prediction in breast cancer. It effectively handles data heterogeneity and identifies molecular signatures defining distinct phenotypic categories. iOmicsPASS treats network

data as undirected graphs, simplifying interaction score calculations, and reduces the number of predictive features. It is suitable for small sample sizes (Koh et al. 2019). NetICS prioritizes cancer genes by integrating diverse data sources on a directed modular interaction network. It identifies mediators between differentially expressed genes, predicting downstream expression changes. Data categories include somatic mutations, gene expressions, DNA copy number variations, methylation patterns, and miRNA expressions. NetICS ranks genes near upstream and downstream differentially expressed genes and aggregates proteins using a robust rank aggregation approach. It outperforms other network-based algorithms in prioritizing cancer genes across multiple cancer types (Dimitrakopoulos et al. 2018).

4. **Similarity based integration:** Similarity-based methods, like SNF (Similarity Network Fusion), build separate networks for each omics data type and iteratively update them to enhance inter-data type similarities, creating a composite network. SNF operates in the sample space, merging mRNA expression, DNA methylation, and miRNA expression data to identify cancer subtypes with distinct survival characteristics (Wang et al. 2014). It's effective for identifying cancer subtypes but not for biomarker discovery. Multiple Kernel Learning (MKL) is ideal for integrating multiple high-throughput data sources, but it's underutilized in genomics due to the lack of standardized protocols and benchmark datasets. An unsupervised version of MKL called Regularized Multiple Kernel Learning Locality Preserving Projections (rMKL-LPP), developed by Speicher and Pfeifer, reduces dimensionality for sample clustering and data analysis. It combines multiple kernel learning with a graph embedding framework algorithm. rMKL-LPP is versatile, suitable for small datasets, and accepts sequence matrices and numerical data as inputs. It aligns with previous clustering results in glioblastoma multiforme, integrating methylation, gene expression, and miRNA expression data (Menyhárt and Győrffy 2021; Speicher and Pfeifer 2015).

5. **Fusion-based integration:** Pattern Fusion Analysis (PFA) is a unique computational technique that uses adaptive optimization to identify integrated "sample-patterns" among diverse genomic profiles. PFA extracts biologically significant sample-patterns in a low-dimensional spatial domain and quantifies the value of each data type or sample in supporting phenotype-specific global sample-patterns. It can identify clinically distinct subgroups in glioblastoma, non-small cell lung cancer, and clear cell carcinoma samples from TCGA with higher prognostic efficacy compared to clustering techniques like iCluster and SNF (Shi et al. 2017). However, PFA doesn't identify new biomarkers or provide insights into tumorigenesis mechanisms.

6. **Correlation-based integration:** Canonical Correlation Analysis (CCA) is a technique for assessing the correlation between gene expression and methylation data, providing insights into tumorigenesis mechanisms. CCA considers individual features while accounting for collective variable impacts (Lin et al. 2013). While useful for estimating survival in cancer, it has limited utility in molecular subtyping and biomarker identification (El-Manzalawy 2018). Computational techniques for biomarker discovery, like MuTarget and matrix-based methods, harness correlations between genetic abnormalities and modifications. MuTarget links somatic mutations and gene expression data, aiding biomarker and therapeutic target discovery (Nagy and Győrffy 2021). DriverNet identifies correlations between mutations and gene expression through influence graphs, revealing interacting gene partners from known pathways (Bashashati et

al. 2012). Multi-Omics Factor Analysis (MOFA) identifies drivers of clinical heterogeneity across multi-omics data, capturing biological and technical variation. It distinguishes shared and unique axes of heterogeneity, enhancing prognostic relevance (Argelaguet et al. 2018). The field of using multi-omics techniques for identifying genetic changes as drivers is rapidly evolving, with cancer hallmark gene lists aiding in connecting driving events to significant signatures (Menyhárt et al. 2016). These approaches offer diverse strategies for multi-data integration and clinical relevance in cancer research.

IV. MULTIOMICS IMPLEMENTATION IN CANCER

The emergence of high-throughput and cost-effective multi-omics approaches, including genomics, epigenomics, transcriptomics, and proteomics, has significantly advanced our understanding of cancer initiation, progression, and treatment. Gradually methods like Cox-nnet, DeepProg, and two-stage Cox-nnet were developed to meet the challenges of integrating the humongous data generated through multi omics approaches (Garmire 2020). These methods have been instrumental in acquiring comprehensive knowledge about cancer at various molecular levels, ultimately leading to more efficient treatment strategies. For instance, in the study by Holowatyj et al. (2020), connections between PPARG visceral adipose tissue expression and plasma/serum markers in colorectal cancer patients were revealed, potentially informing therapeutic strategies. In diseases like prostate adenocarcinoma, which require a multi-omics approach, researchers have identified tumor suppressor genes and gene mutations. PCDH9, a gene absent in 23% of cases, has been associated with other molecular changes, providing valuable insights into prognosis (Ren et al. 2016). Multi-omics analyses have been instrumental in numerous cancer studies, including melanoma (Zhang et al. 2020), breast cancer (Sandri et al. 2018), pancreatic ductal adenocarcinoma (Chaudhary et al. 2018), and more. These studies have identified predictive biomarkers, offered insights into disease progression, and explored potential therapeutic avenues (Kwon et al. 2015, Zhu et al. 2017, Yoshikawa et al. 2017). While single-level omics approaches have provided valuable insights into cancer biology, multi-omics analyses are indispensable for understanding the complex interactions between molecular variations and phenotypic manifestations. Multi-omics data integration allows for a comprehensive view of genetic variants, environmental factors, and intricate biological system interactions, contributing to prognostic and predictive studies, therapy response investigations, and translational research (Menyhárt and Győrffy 2021). Integrative multi-omics data provide a detailed perspective on tumorigenesis, enabling better patient selection for targeted therapies and the optimization of clinical treatment strategies. As technological barriers are overcome, multi-omics techniques continue to advance cancer research, offering substantial benefits to cancer patients worldwide.

V. CHALLENGES IN INTEGRATED OMICS

Integrating a wide range of omics data poses both conceptual and practical challenges in day-to-day omics data analysis. Key challenges in current integrated omics techniques include sample preparation, normalization, transformation of multiple omics datasets, integration issues, data archiving and sharing, data interpretation, and the complexities of clinical translation in multiomics approaches. Numerous studies have emphasized the complexities of proficiently preparing samples and processing data from diverse sources in individual omics research, spanning microorganisms, plants, and animals (Misra et al., 2019).

With multi-omics, sample size becomes a critical constraint, further exacerbated by the need for unified extraction techniques capable of simultaneously extracting proteins, nucleic acids, and metabolites without significant loss. Each omics platform has its unique normalization and transformation techniques due to the diversity of information within datasets. For example, a zero value is interpreted differently in different omics datasets, representing non-expression in RNA-Seq-based transcriptome data but merely missing data in proteomic or metabolomic datasets. In many cases, multiple omics datasets may not fully overlap, and measures derived from one omics method may not be well correlated with those from other approaches. Consequently, individual omics techniques provide an incomplete picture of the complex pathophysiology of complex diseases. Additionally, combining data from multiple sources complicates the identification of false positives within integrated datasets, requiring careful consideration of approaches to address this issue. Effective data archiving is crucial for the robustness of individual omics and integrated omics data, aligning with the principles of Findability, Accessibility, Interoperability, and Reusability (FAIR) (Wilkinson et al., 2016). While numerous public databases exist for archiving individual omics datasets, there is a lack of standardized databases that enable users to submit and retrieve integrated omics data from a unified repository or interface. Sharing data, especially in large multi-omics studies, can enhance resource accessibility for further training, exploration, and post-publication analyses. Interpreting and managing vast, multifaceted networks is computationally and time-intensive, demanding a deep understanding of the biological system under investigation. Despite the growing number of potential cancer biomarkers discovered across various omics levels, the development of new cancer biomarker- based diagnostics has been limited. The translational process faces several challenges, contributing to the gap between biomarker discovery and clinical adoption (Maes et al., 2015). To translate biomarker discovery into clinical assays, collaborative efforts among academic researchers, clinicians, and industry experts are essential to establish clinical significance, validation procedures, and study designs. Building such interdisciplinary teams can be resource- intensive and remains a significant hurdle in translational research transformation techniques. For instance, a zero value is treated differently in different omics datasets. It is intercepted as non-expression for the transcript in RNA-Seq-based transcriptome dataset while it can be merely missing data for proteomic or metabolomic dataset. Multiple omics datasets may not always overlap, and measures derived from one omics method aren't necessarily well associated with those derived from other approaches. As a result, different omics techniques individually give an incomplete data of the complicated pathophysiology of complex diseases. Furthermore, combining data from several sources makes it more difficult to account for false positives in the integrated datasets. The approach to deal with false positives in different omics datasets has a major impact on the outcomes. Data archiving is very essential for robustness of individual omics and integrated omics data, including adherence to Findability, Accessibility, Interoperability and Reusability (FAIR) principles (Wilkinson et al. 2016). Although many public databases are available for archiving individual omics datasets, there are no standard databases which let users to submit and retrieve three or more integrated (López de Maturana et al. 2019).

VI. CONCLUSION

Despite extensive research, only a handful of molecular diagnostic tests have made significant progress towards clinical use since their discovery a decade ago, while many others have been deemed research failures. To address these disparities, there is a pressing need for robust regulatory standards to identify, assess, and validate these biomolecules.

Once validated, these biomolecules can function as diagnostic, prognostic, predictive, and therapeutic markers, significantly impacting patient outcomes through early tumor detection, tailored targeted therapies, and personalized treatment approaches.

VII. ACKNOWLEDGEMENTS

Department of Science and Technology, Government of Odisha (Odisha University Research and Innovation Incentivization Plan Grant No. 590/83/OSHEC); University Grants Commission, Government of India (NTA Ref. No.: 211610122720); Higher Education Department, Government of Odisha (Grant No 26913/HED/HE□PTC□WB□02□17(OHEPEE); Indian Council of Medical Research (ICMR) (Grant No.: 2021-10784/CMB-BMS).

1. **Authorship conformation or contribution statement:** Conceptualization: Luna Samanta. Data curation: Amrita Swain, Anwesha Pradhan and Soumya Ranjan Jena. Writing-original draft preparation: Amrita Swain, Anwesha Pradhan, Soumya Ranjan Jena. Writing, review & editing: Luna Samanta and Dhananjay Soren. Supervision: LunaSamanta and Dhananjay Soren. Funding acquisition: Luna Samanta and Dhananjay Soren. All authors read and approved the final manuscript.

2. **Conflict of Interest:** The authors declare that there is no conflict of interest.

3. **Funding statement:** Department of Science and Technology, Government of India (SERB Grant No. EEQ/2016/000808); Higher Education Department, Government of Odisha (Grant No 26913/HED/HE□PTC□WB□02□17(OHEPEE).

4. **Figure to Legends**

- **Figure 1:** Schematic representation of work flow of different omics platforms in cancer research for biomarker discovery.
- **Figure 2:** Schematic representation of methods used for multi-omics data integration showcasing work flow for biomarker discovery in cancer.

REFERENCES

[1] Alessandro R, Belluco C, Kohn EC. Proteomic approaches in colon cancer: promising tools for new cancer markers and drug target discovery. Clinical colorectal cancer. 2005;4(6):396-402.
[2] Alyass A, Turcotte M, Meyre D. From big data analysis to personalized medicine for all: challenges and opportunities. BMC medical genomics. 2015;8(1):1-12.
[3] Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, et al. Multi□Omics Factor Analysis a framework for unsupervised integration of multi□omics data sets. Molecular systems biology. 2018;14(6):e8124.
[4] Bashashati A, Haffari G, Ding J, Ha G, Lui K, Rosner J, et al. DriverNet: uncovering the impact of somatic driver mutations on transcriptional networks in cancer. Genome biology. 2012;13(12):1-14.
[5] Bayón GF, Fernández AF, Fraga MF. Bioinformatics Tools in Epigenomics Studies. Epigenomics in Health and Disease: Elsevier; 2016. p. 73-107.
[6] Bonnet E, Calzone L, Michoel T. Integrative multi-omics module network inference with Lemon-Tree. PLoS computational biology. 2015;11(2):e1003983.
[7] Chakraborty S, Hosen M, Ahmed M, Shekhar HU. Onco-multi-OMICS approach: a new frontier in cancer research. BioMed research international. 2018;2018.
[8] Champion M, Brennan K, Croonenborghs T, Gentles AJ, Pochet N, Gevaert O. Module analysis captures

pancancer genetically and epigenetically deregulated cancer driver genes for smoking and antiviral response. EBioMedicine. 2018;27:156-66.
[9] Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep learning–based multi-omics integration robustly predicts survival in liver cancer. Clinical Cancer Research. 2018;24(6):1248-59.
[10] Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. Genome biology. 2016;17(1):1-19.
[11] Conley BA, Taube SE. Prognostic and predictive markers in cancer. Disease markers. 2004;20(2):35-43.
[12] Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012;486(7403):346-52.
[13] Dalal N, Jalandra R, Sharma M, Prakash H, Makharia GK, Solanki PR, et al. Omics technologies for improved diagnosis and treatment of colorectal cancer: Technical advancement and major perspectives. Biomedicine & Pharmacotherapy. 2020;131:110648.
[14] Dhingra P, Martinez-Fundichely A, Berger A, Huang FW, Forbes AN, Liu EM, et al. Identification of novel prostate cancer drivers using RegNetDriver: a framework for integration of genetic and epigenetic alterations with tissue-specific regulatory network. Genome biology. 2017;18(1):1-23.
[15] Dimitrakopoulos C, Hindupur SK, Häfliger L, Behr J, Montazeri H, Hall MN, et al. Network-based integration of multi-omics data for prioritizing cancer genes. Bioinformatics. 2018;34(14):2441-8.
[16] Dorcely B, Katz K, Jagannathan R, Chiang SS, Oluwadare B, Goldberg IJ, et al. Novel biomarkers for prediabetes, diabetes, and associated complications. Diabetes, metabolic syndrome and obesity: targets and therapy. 2017;10:345.
[17] El-Manzalawy Y, editor CCA based multi-view feature selection for multi-omics data integration. 2018 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB); 2018: IEEE.
[18] Fagerquist CK, Sultan O. Induction and identification of disulfide-intact and disulfide-reduced β-subunit of Shiga toxin 2 from Escherichia coli O157: H7 using MALDI-TOF-TOF-MS/MS and top-down proteomics. Analyst. 2011;136(8):1739- 46.
[19] Fernandez-Lozano C, Blanco JL, Gestal M, Dorado J, Pazos A, editors. Integrative multi-omics data-driven approach for metastasis prediction in cancer. Proceedings of the First International Conference on Data Science, E-learning and Information Systems; 2018.
[20] Fujiwara H, Suzuki M, Takeshima N, Takizawa K, Kimura E, Nakanishi T, et al. Evaluation of human epididymis protein 4 (HE4) and Risk of Ovarian Malignancy Algorithm (ROMA) as diagnostic tools of type I and type II epithelial ovarian cancer in Japanese women. Tumor Biology. 2015;36(2):1045-53.
[21] Garmire LX. Strategies to integrate multi-omics data for patient survival prediction. arXiv preprint arXiv:200812455. 2020.
[22] Geng X, Ke C, Chen G, Liu P, Wang F, Zhang H, et al. On-line separation of native proteins by two-dimensional liquid chromatography using a single column. Journal of Chromatography A. 2009;1216(16):3553-62.
[23] Goossens N, Nakagawa S, Sun X, Hoshida Y. Cancer biomarker discovery and validation. Transl Cancer Res 4 (3): 256–269. 2015.
[24] Holowatyj AN, Haffa M, Lin T, Scherer D, Gigic B, Ose J, et al. Multi-omics Analysis Reveals Adipose–tumor Crosstalk in Patients with Colorectal Cancer. Cancer Prevention Research. 2020;13(10):817-28.
[25] Huang S, Chaudhary K, Garmire LX. More is better: recent progress in multi-omics data integration methods. Frontiers in genetics. 2017;8:84.
[26] Itkonen HM, Poulose N, Walker S, Mills IG. CDK9 inhibition induces a metabolic switch that renders prostate cancer cells dependent on fatty acid oxidation. Neoplasia. 2019;21(7):713-20.
[27] Kamoun A, Cancel-Tassin G, Fromont G, Elarouci N, Armenoult L, Ayadi M, et al. Comprehensive molecular classification of localized prostate adenocarcinoma reveals a tumour subtype predictive of non-aggressive disease. Annals of Oncology. 2018;29(8):1814-21.
[28] Kiebish MA, Cullen J, Mishra P, Ali A, Milliman E, Rodrigues LO, et al. Multi-omic serum biomarkers for prognosis of disease progression in prostate cancer. Journal of translational medicine. 2020;18(1):1-10.
[29] Koh HW, Fermin D, Vogel C, Choi KP, Ewing RM, Choi H. iOmicsPASS: network- based integration of multiomics data for predictive subnetwork discovery. NPJ systems biology and applications. 2019;5(1):1-10.
[30] Kori M, Yalcin Arga K. Potential biomarkers and therapeutic targets in cervical cancer: Insights from the meta-analysis of transcriptomics data within network biomedicine perspective. PloS one. 2018;13(7):e0200717.
[31] Kwon M-S, Kim Y, Lee S, Namkung J, Yun T, Yi SG, et al. Integrative analysis of multi-omics data for

identifying multi-markers for diagnosing pancreatic cancer. BMC genomics. 2015;16(9):1-10.
[32] Kwon OK, Jeon JM, Sung E, Na A-Y, Kim SJ, Lee S. Comparative secretome profiling and mutant protein identification in metastatic prostate cancer cells by quantitative mass spectrometry-based proteomics. Cancer genomics & proteomics. 2018;15(4):279-90.
[33] Li J-Y, Li C-J, Lin L-T, Tsui K-H. Multi-omics analysis identifying key biomarkers in ovarian cancer. Cancer Control. 2020;27(1):1073274820976671.
[34] Li L, Wei Y, To C, Zhu C-Q, Tong J, Pham N-A, et al. Integrated omic analysis of lung cancer reveals metabolism proteome signatures with prognostic impact. Nature communications. 2014;5(1):1-12.
[35] Li Z, Gao X, Peng X, Chen M-JM, Li Z, Wei B, et al. Multi-omics characterization of molecular features of gastric cancer correlated with response to neoadjuvant chemotherapy. Science advances. 2020;6(9):eaay4211.
[36] Lin D, Zhang J, Li J, Calhoun VD, Deng H-W, Wang Y-P. Group sparse canonical correlation analysis for genomic data integration. BMC bioinformatics. 2013;14(1):1- 16.
[37] Lock EF, Dunson DB. Bayesian consensus clustering. Bioinformatics. 2013;29(20):2610-6.
[38] Lock EF, Hoadley KA, Marron JS, Nobel AB. Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. The annals of applied statistics. 2013;7(1):523.
[39] López de Maturana E, Alonso L, Alarcón P, Martín-Antoniano IA, Pineda S, Piorno L, et al. Challenges in the integration of omics and non-omics data. Genes. 2019;10(3):238.
[40] Lu M, Zhan X. The crucial role of multiomic approach in cancer research and clinically relevant outcomes. EPMA Journal. 2018;9(1):77-102.
[41] Ma Z-Q, Tabb DL, Burden J, Chambers MC, Cox MB, Cantrell MJ, et al. Supporting tool suite for production proteomics. Bioinformatics. 2011;27(22):3214-5.
[42] Maes E, Cho WC, Baggerman G. Translating clinical proteomics: the importance of study design. Expert review of proteomics. 2015;12(3):217-9.
[43] Mallik S, Sen S, Maulik U. IDPT: Insights into potential intrinsically disordered proteins through transcriptomic analysis of genes for prostate carcinoma epigenetic data. Gene. 2016;586(1):87-96.
[44] Mao Y, Valeja SG, Rouse JC, Hendrickson CL, Marshall AG. Top-down structural analysis of an intact monoclonal antibody by electron capture dissociation-Fourier transform ion cyclotron resonance-mass spectrometry. Analytical chemistry. 2013;85(9):4239-46.
[45] Maruvada P, Wang W, Wagner PD, Srivastava S. Biomarkers in molecular medicine: cancer detection and diagnosis. Biotechniques. 2005;38(S4):S9-S15.
[46] Masica DL, Karchin R. Correlation of somatic mutation and expression identifies genes important in human glioblastoma progression and survival. Cancer research. 2011;71(13):4550-61.
[47] Meng C, Helm D, Frejno M, Kuster B. moCluster: identifying joint patterns across multiple omics data sets. Journal of proteome research. 2016;15(3):755-65.
[48] Menyhárt O, Győrffy B. Multi-omics approaches in cancer research with applications in tumor subtyping, prognosis, and diagnosis. Computational and Structural Biotechnology Journal. 2021;19:949.
[49] Menyhárt O, Harami-Papp H, Sukumar S, Schäfer R, Magnani L, de Barrios O, et al. Guidelines for the selection of functional assays to evaluate the hallmarks of cancer. Biochimica et Biophysica Acta (BBA)-Reviews on Cancer. 2016;1866(2):300-19.
[50] Mertins P, Mani D, 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.
[51] Michaut M, Chin S-F, Majewski I, Severson TM, Bismeijer T, De Koning L, et al. Integration of genomic, transcriptomic and proteomic data identifies two biologically distinct subtypes of invasive lobular breast cancer. Scientific reports. 2016;6(1):1-13.
[52] Misra BB, Langefeld C, Olivier M, Cox LA. Integrated omics: tools, advances and future approaches. Journal of molecular endocrinology. 2019;62(1):R21-R45.
[53] Mo Q, Shen R, Guo C, Vannucci M, Chan KS, Hilsenbeck SG. A fully Bayesian latent variable model for integrative clustering analysis of multi-type omics data. Biostatistics. 2018;19(1):71-86.
[54] Mo Q, Wang S, Seshan VE, Olshen AB, Schultz N, Sander C, et al. Pattern discovery and cancer gene identification in integrated cancer genomic data. Proceedings of the National Academy of Sciences. 2013;110(11):4245-50.
[55] Murphy K, Murphy BT, Boyce S, Flynn L, Gilgunn S, O'Rourke CJ, et al. Integrating biomarkers across omic platforms: an approach to improve stratification of patients with indolent and aggressive prostate cancer. Molecular oncology. 2018;12(9):1513- 25.
[56] Nagy Á, Győrffy B. muTarget: A platform linking gene expression changes and mutation status in solid tumors. International journal of cancer. 2021;148(2):502-11.
[57] Nam H, Chung BC, Kim Y, Lee K, Lee D. Combining tissue transcriptomics and urine metabolomics for

breast cancer biomarker identification. Bioinformatics. 2009;25(23):3151-7.
[58] Rappoport N, Shamir R. Multi-omic and multi-view clustering algorithms: review and cancer benchmark. Nucleic acids research. 2018;46(20):10546-62.
[59] Ray P, Zheng L, Lucas J, Carin L. Bayesian joint analysis of heterogeneous genomics data. Bioinformatics. 2014;30(10):1370-6.
[60] Ren S, Shao Y, Zhao X, Hong CS, Wang F, Lu X, et al. Integration of metabolomics and transcriptomics reveals major metabolic pathways and potential biomarker involved in prostate cancer. Molecular & Cellular Proteomics. 2016;15(1):154-63.
[61] Roberts NJ, Vogelstein JT, Parmigiani G, Kinzler KW, Vogelstein B, Velculescu VE. The predictive capacity of personal genome sequencing. Science translational medicine. 2012;4(133):133ra58-ra58.
[62] Sandri BJ, Kaplan A, Hodgson SW, Peterson M, Avdulov S, Higgins L, et al. Multi- omic molecular profiling of lung cancer in COPD. European Respiratory Journal. 2018;52(1).
[63] Shen R, Olshen AB, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. Bioinformatics. 2009;25(22):2906-12.
[64] Shi Q, Zhang C, Peng M, Yu X, Zeng T, Liu J, et al. Pattern fusion analysis by adaptive alignment of multiple heterogeneous omics data. Bioinformatics. 2017;33(17):2706-14.
[65] Song X, Ji J, Gleason KJ, Yang F, Martignetti JA, Chen LS, et al. Insights into impact of DNA copy number alteration and methylation on the proteogenomic landscape of human ovarian cancer via a multi-omics integrative analysis. Molecular & Cellular Proteomics. 2019;18(8):S52-S65.
[66] Speicher NK, Pfeifer N. Integrating different data types by regularized unsupervised multiple kernel learning with application to cancer subtype discovery. Bioinformatics. 2015;31(12):i268-i75.
[67] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2021;71(3):209- 49.
[68] Vasaikar S, Huang C, Wang X, Petyuk VA, Savage SR, Wen B, et al. Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. Cell. 2019;177(4):1035-49. e19.
[69] 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. Nature methods. 2014;11(3):333-7.
[70] Wang KC, Chang HY. Epigenomics: technologies and applications. Circulation research. 2018;122(9):1191-9.
[71] Wang W, Baladandayuthapani V, Morris JS, Broom BM, Manyam G, Do K-A. iBAG: integrative Bayesian analysis of high-dimensional multiplatform genomics data. Bioinformatics. 2013;29(2):149-59.
[72] Wang Z, Wei Y, Zhang R, Su L, Gogarten SM, Liu G, et al. Multi-omics analysis reveals a HIF network and hub gene EPAS1 associated with lung adenocarcinoma. EBioMedicine. 2018;32:93-101.
[73] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data. 2016;3(1):1-9.
[74] Woo HG, Choi J-H, Yoon S, Jee BA, Cho EJ, Lee J-H, et al. Integrative analysis of genomic and epigenomic regulation of the transcriptome in liver cancer. Nature communications. 2017;8(1):1-11.
[75] Wu L, Qu X. Cancer biomarker detection: recent achievements and challenges. Chemical Society Reviews. 2015;44(10):2963-97.
[76] Wu L, Yang Y, Guo X, Shu X-O, Cai Q, Shu X, et al. An integrative multi-omics analysis to identify candidate DNA methylation biomarkers related to prostate cancer risk. Nature communications. 2020;11(1):1-11.
[77] Yan S-K, Liu R-H, Jin H-Z, Liu X-R, Ye J, Shan L, et al. “Omics” in pharmaceutical research: overview, applications, challenges, and future perspectives. Chinese journal of natural medicines. 2015;13(1):3-21.
[78] Yang K, Xia B, Wang W, Cheng J, Yin M, Xie H, et al. A comprehensive analysis of metabolomics and transcriptomics in cervical cancer. Scientific reports. 2017;7(1):1- 11.
[79] Yang Z, Liu B, Lin T, Zhang Y, Zhang L, Wang M. Multiomics analysis on DNA methylation and the expression of both messenger RNA and microRNA in lung adenocarcinoma. Journal of cellular physiology. 2019;234(5):7579-86.
[80] Yoshikawa S, Kiyohara Y, Otsuka M, Kondou R, Nonomura C, Miyata H, et al. Multi-omics profiling of patients with melanoma treated with Nivolumab in project HOPE. Anticancer research. 2017;37(3):1321-8.
[81] Yousefi M, Rajaie S, Keyvani V, Bolandi S, Hasanzadeh M, Pasdar A. Clinical significance of circulating tumor cell related markers in patients with epithelial ovarian cancer before and after adjuvant chemotherapy. Scientific reports. 2021;11(1):1-10.
[82] Yu X-T, Zeng T. Integrative analysis of omics big data. Computational Systems Biology: Humana Press,

New York, NY; 2018. p. 109-35.
[83] Yurgelun MB, Hampel H. Recent advances in lynch syndrome: diagnosis, treatment, and cancer prevention. American Societyof Clinical Oncology Educational Book. 2018;38:101-9.
[84] Zhang B, Wang J, Wang X, Zhu J, Liu Q, Shi Z, et al. Proteogenomic characterization of human colon and rectal cancer. Nature. 2014;513(7518):382-7.
[85] Zhang E, Zhang M, Shi C, Sun L, Shan L, Zhang H, et al. An overview of advances in multi-omics analysis in prostate cancer. Life sciences. 2020:118376.
[86] 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.
[87] Zhang Q, Lou Y, Yang J, Wang J, Feng J, Zhao Y, et al. Integrated multiomic analysis reveals comprehensive tumour heterogeneity and novel immunophenotypic classification in hepatocellular carcinomas. Gut. 2019;68(11).
[88] Zhang S, Liu C-C, Li W, Shen H, Laird PW, Zhou XJ. Discovery of multi- dimensional modules by integrative analysis of cancer genomic data. Nucleic acids research. 2012;40(19):9379-91.
[89] Zhu B, Song N, Shen R, Arora A, Machiela MJ, Song L, et al. Integrating clinical and multiple omics data for prognostic assessment across human cancers. Scientific reports. 2017;7(1):1-13.

