Computational and Structural Biotechnology Journal 23 (2024) 954–971

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

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

Mini-review

A review on trends in development and translation of omics signatures in cancer

Wei Ma [a,1], Wenshu Tang [a,1], Jamie S.L. Kwok [a], Amy H.Y. Tong [a], Cario W.S. Lo [a], Annie T.W. Chu [a], Brian H.Y. Chung [a,b,*], Hong Kong Genome Project

[a] Hong Kong Genome Institute, Hong Kong, China
[b] Department of Pediatrics and Adolescent Medicine, School of Clinical Medicine, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, China

ARTICLE INFO

Keywords:
Cancer genomics
Transcriptomics
DNA methylation
Mutational signatures
Machine learning
Translational Medicine

ABSTRACT

The field of cancer genomics and transcriptomics has evolved from targeted profiling to swift sequencing of individual tumor genome and transcriptome. The steady growth in genome, epigenome, and transcriptome datasets on a genome-wide scale has significantly increased our capability in capturing signatures that represent both the intrinsic and extrinsic biological features of tumors. These biological differences can help in precise molecular subtyping of cancer, predicting tumor progression, metastatic potential, and resistance to therapeutic agents. In this review, we summarized the current development of genomic, methylomic, transcriptomic, proteomic and metabolic signatures in the field of cancer research and highlighted their potentials in clinical applications to improve diagnosis, prognosis, and treatment decision in cancer patients.

1. Introduction

With the advance in sequencing technologies, whole-genome and whole-transcriptome sequencing have now been widely used in cancer studies to provide researchers with information of DNA mutation, DNA methylation and gene expression at a genome-wide level. Together with the advances in mass spectrometry which enabled in depth analysis of proteomic and metabolic profiles in cancers, the omics signatures have been studied for their use in different clinical applications in cancer patients (Fig. 1). It has been shown that by discerning generic patterns in genomic data, one can reveal the basic biological properties of a tumor such as defects in DNA-repair pathways [1]. For example, consistent patterns of DNA mutagenesis across breast tumours can be used to identify *BRCA1*-null and *BRCA2*-null tumors [2]. In some cases, one can even distinguish tumors showing impairment in the *BRCA*-pathway but without having apparent mutations in the *BRCA1/2* genes themselves, which enables identification of patients who will respond better to PARP inhibitors [3]. This lifts off the limitation of our knowledge of biology from preventing clinical translation of genomic and transcriptomic results. This is, in part, due to the fact that genome-wide signature analysis of DNA mutation, DNA methylation and gene expression can reveal not only the effect of tumor cell intrinsic pathways but also the interaction between tumor cells and the microenvironment [4–6]. Moreover, development of new mathematical and computational tool is essential in improving the understanding of the complex genome-wide sequencing data and extracting abstract patterns which are associated with specific biological and/or clinical features of the tumor [7–9].

2. DNA mutational signature in cancer

Mutations accumulate with age and multiple exposures, engraving characteristic mutational patterns or imprints in the genome of somatic cells, termed as mutational signatures. Mutational signatures represent genome-wide somatic alteration patterns and reflect the activities of endogenous and exogenous mutational processes. For example, DNA repair deficiencies can leave specific footprints on the cancer genome. Therefore, depicting mutational signatures provides a conceptual breakthrough to understanding life history and tumor etiology at the DNA level and underpins advances in cancer early detection and therapeutic implications.

2.1. Detecting SNV mutational signatures

Since the initial identification of merely five SNV signatures in a

* Correspondence to: Hong Kong Genome Institute, 2/F, Building 20E, Hong Kong Science Park, Hong Kong, China.
E-mail address: bhychung@genomics.org.hk (B.H.Y. Chung).
[1] Equal contribution

https://doi.org/10.1016/j.csbj.2024.01.024
Received 27 October 2023; Received in revised form 31 January 2024; Accepted 31 January 2024
Available online 3 February 2024

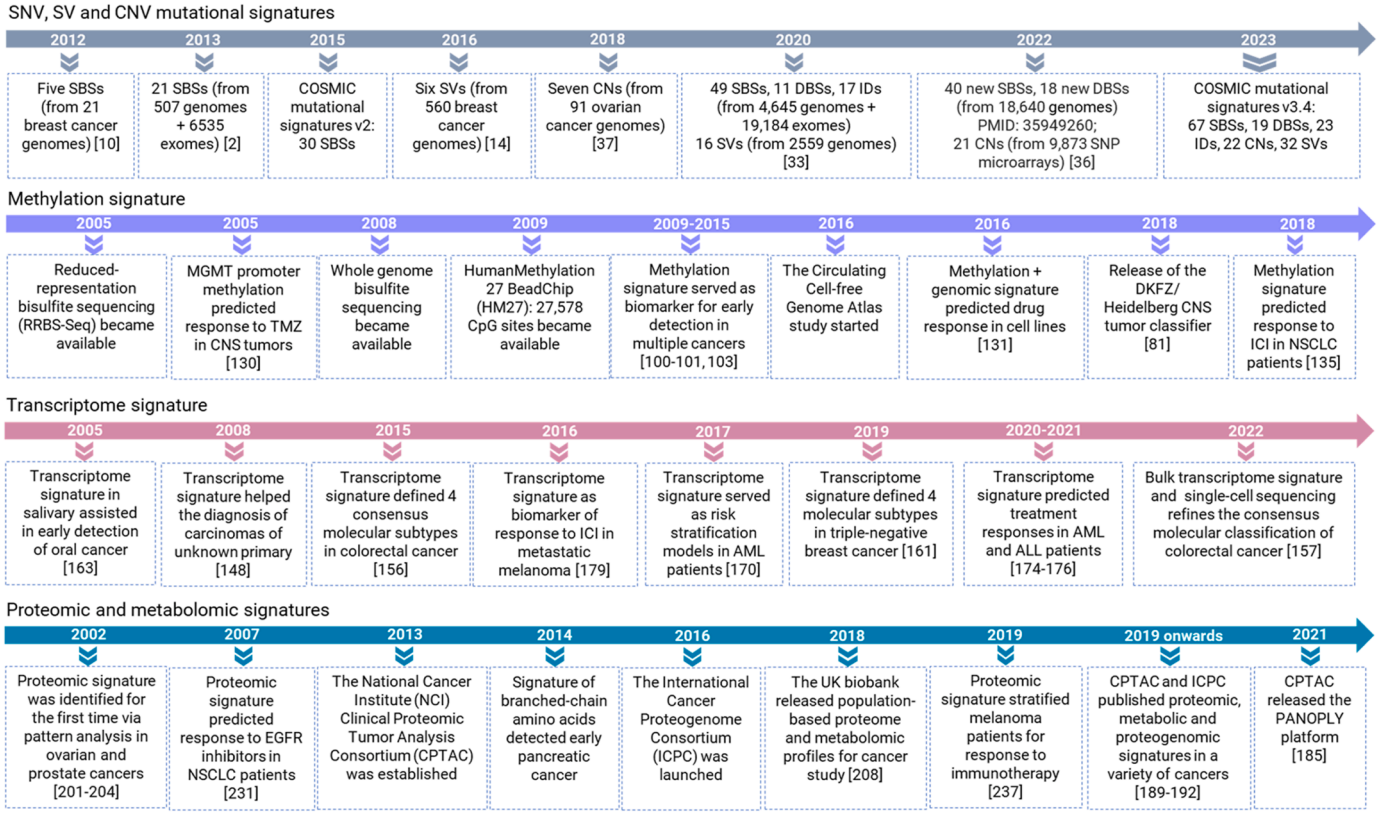

Fig. 1. Timeline of milestones of different omics signatures. Fig. 1. The figure shows some of the key milestones for 1. Mutational signatures, 2. Methylation signature, 3. Transcriptome signature and 4. Proteomic and metabolic signatures. Abbreviation: SNV, single-nucleotide variant; SV, structural variation; CNV, copy number variation; SBS, single base substitution; DBS, double base substitution; ID, insertions and deletion; TMZ, temozolomide; DKFZ, The German Cancer Research Center; AML, acute myeloid leukemia; ALL, acute lymphocytic leukemia; NSCLC, non-small cell lung cancer.

study of 21 breast cancer cases back in 2012 [10], there has been a significant expansion in the catalog of SNV signatures (Fig. 2). Large-scale genome projects such as the Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium of the International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA) have uncovered an increasingly complex molecular landscape of human cancers [11,12]. These projects have amalgamated genomic data from an impressive collection of over 4000 whole genomes and nearly 20,000 exomes. This extensive integration has paved the way for the identification of a multitude of signatures that encompass many cancer types, which are now deposited in the Catalogue of Somatic Mutations in Cancer (COSMIC) (https://cancer.sanger.ac.uk/cosmic) [2,13–15]. Advancing further, the analysis of whole genome sequences has remarkably facilitated the identification of novel mutational signatures [16].

The etiology of certain mutational signatures has been successfully validated. Some of these signatures are associated with known exogenous mutagens such as ultraviolet radiation and smoking. Other signatures are associated with endogenous exposures such as replication errors arising from impaired DNA repair processes [17], homologous recombination deficiency (HRd) [18], mismatch repair deficiency (MMRd) [19] and nucleotide excision repair deficiency (NERd) [20].

SNV mutational signatures fall into three categories: single base substitution (SBS), double base substitution (DBS), and small insertions and deletion (ID). The concept of SBS and DBS were introduced based on a distinct type of one or two consecutive nucleotide substitution(s) and combination of 5' and 3' neighboring bases, resulting in 96/78 possible mutation types, respectively. The overall pattern of the 96/78 channels of mutation is the mutational signature. The progression of sequencing techniques has spurred the creation of computational methods for deducing mutational signatures from individual samples [16,21–24]. Alexandrov et al. pioneered inference of mutational signatures by developing a non-negative matrix factorization (NMF)-based WTSI framework [Wellcome Trust Sanger Institute (WTSI) Mutational Signature Framework] for *de novo* signature extraction [21]. This approach allows for the unbiased identification of both predominantly and novel signatures and is predominantly used for deriving reference signatures from extensive cohorts. The principle involves decomposing a matrix M (somatic mutations x samples) into a matrix S (a set of mutational signature references) and a matrix A (activity of each signature) to determine the signature profiles and contributions of each signature in each cancer genome. From then on, a variety of NMF-based tools have been developed for identifying mutational signatures, including *Maftools, MutationalPatterns*, *MutSpec*, *SignatureToolsLib*, *SigMiner*, *SomaticSignatures*, and *SigProfiler_PCAWG* [15,22–27]. For example, *SigProfiler_PCAWG*, an enhanced version of the original WTSI framework, was employed to construct the reference signatures deposited in the COSMIC database through an unsupervised machine-learning approach for *de novo* extraction of signatures [15]. *SigProfilerExtractor*, a more recent version that implements a PyTorch-based factorization method, automatically selects the number of signatures, and decomposes *de novo* extracted signatures to known COSMIC signatures [28].

Apart from *de novo* extraction, re-fitting is another approach that allows the assignment of a predefined set of signatures (i.e., COSMIC Legacy SBS and COSMIC V3 SBS, DBS and ID) to an individual sample for signature identification. Several signature refitting tools have been developed, such as *deconstructSigs*, *mutationalPatterns,* and *SigProfilerAssignment* [23,29,30]. Furthermore, supervised machine learning (ML) tools that incorporate known exposures during the training of the algorithm have emerged, such as *SuperSigs*. *SuperSigs* considers clinical factors such as age, smoking status, and body mass index, enhancing its

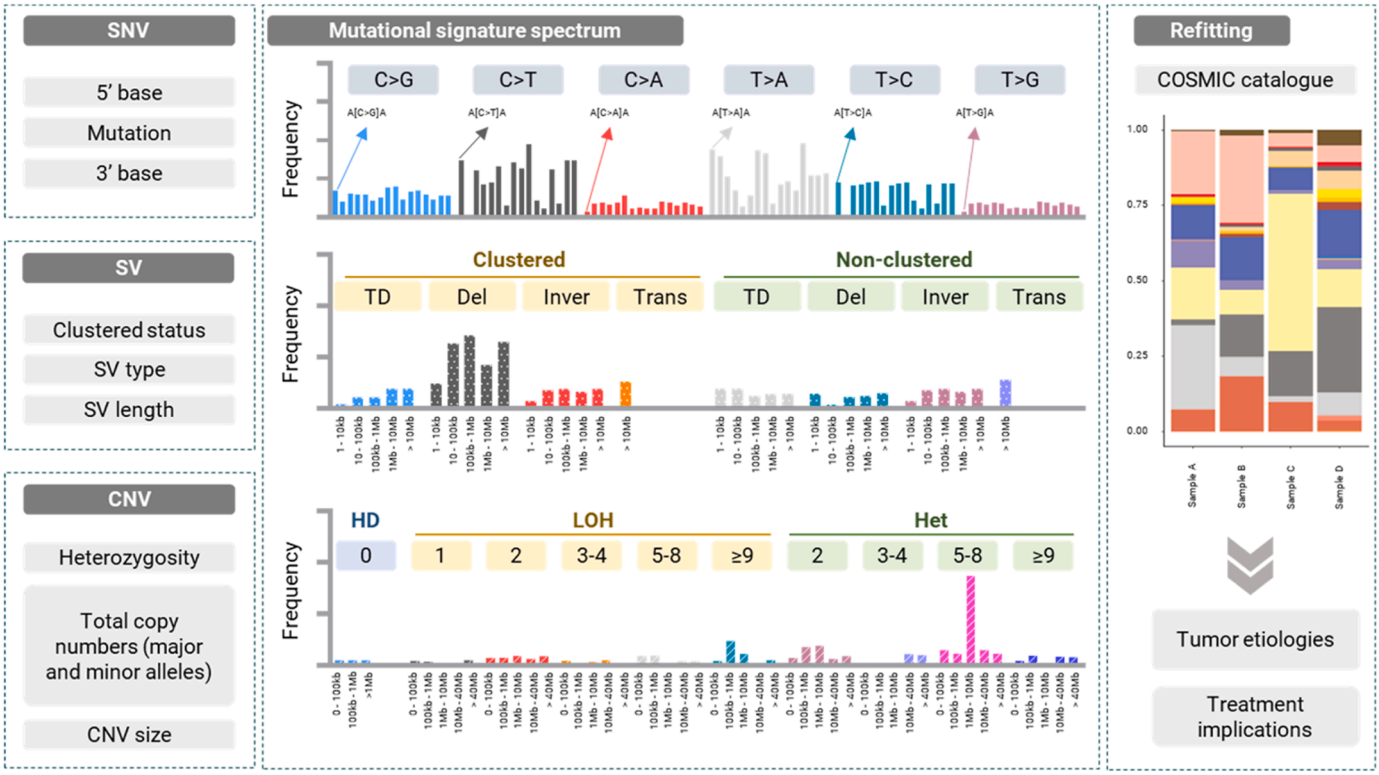

Fig. 2. Summary of mutational signatures based on SNV, CNV and SV. Fig. 2. The figure depicts features of SNV, SV and CNV mutational signatures. For SNVs, the 5' and 3' of the mutated site, as well as the type of mutation are considered. For SVs, four types of variants, the distance between the two breakpoints, along with the clustered status of the SVs are considered. For CNVs, factors such as heterozygosity, total copy number, and CNV size are taken into consideration. Mutational signature fitting is a mathematical procedure used to determine the combination of known signatures, such as the COSMIC catalogue. Abbreviation: SNV, single-nucleotide variant; SV, structural variation; CNV, copy number variation; HD, homozygous deletion; LOH, loss of heterozygosity; Het, heterozygous.

effectiveness in predicting environmental or lifestyle factors that may contribute to cancer. This method has been successful in identifying mutation patterns linked to cancer in obese patients [31].

While fitting methods offer a more precise estimation of each signature's relative and absolute contribution to each sample, overfitting is a concern that warrants careful consideration. Maura et al. have proposed a multi-step guideline for a more accurate mutational signature analysis in hematological cancers. This involves initially shortlisting COSMIC signatures using *de novo* extraction methods, followed by refitting each individual sample with only the subset of COSMIC signatures identified earlier [32].

2.2. Detecting SV/CNV mutational signatures

Structural variations (SVs) are prevalent in cancer, defined by large-scale intra-chromosomal events (>50 bp in size) or inter-chromosomal events. Classification schemas for SVs have also been actively explored to better understand the mutagenic processes of the tumor [32,33]. SV signatures were first described in breast cancer, among which three SV signatures characterized by tandem duplications or deletions were identified to be associated with HRd [14]. Recently, a pan-cancer study has investigated SV signatures by integrating both canonical and complex SV such as chromothripsis, chromoplexy, multiple inversions and templated insertion. Notably, SV signatures consistently correlated with the presence of pathogenic germline variants and somatic driver mutations in DNA-repair genes, including *BRCA1*, *BRCA2*, *FANC*, *CDK12*, and *PALB2* across a wide spectrum of human cancers [33]. Maclachlan et al. has used this classification approach to investigate SV signatures in multiple myeloma (MM) patients and found SV signatures associated with distinct and known MM drivers, which helps understand the genomic complexity in MM [34]. Moreover, a recent study has identified complex rearrangements by analyzing the topology of junction copy numbers of the rearrangement. Clustering of tumors according to the complex rearrangements identified subgroups associated with DNA repair defects and poor prognosis [35]. The first set of SV signatures was recently released in COSMIC v3.4. This set includes 10 types of SVs, derived from an analysis of 10,731 whole genomes across 16 different tissue types provided by Genomics England.

Copy number variations, referring to gains and losses of DNA, belong to an important class of somatic mutation and emerge because of errors in replication, mitotic recombination, and breakage-fusion-bridge cycles, which may lead to chromosomal instability [36]. The pattern of pan-cancer CNV signatures was first examined by analyzing global balanced or unbalanced CNV events such as whole-genome doubling, aneuploidy and loss of heterozygosity. As a result, 21 pan-cancer CNV signatures associated with different etiologies including chromothripsis, LOH, HRd and cancer-driver genes have been identified, among which the chromothripsis-related CNV signature is associated with poor disease-specific survival in glioblastoma [36]. Macintyre et al. demonstrated that copy number signatures predict overall survival and the probability of platinum-resistant relapse in ovarian cancer [37]. Drews et al. have developed an analytical framework, through which they identified 17 distinct signatures that are associated with specific types of chromosomal instability, leading to substantial DNA alteration events within the cancer genome [38].

The identification of canonical SVs patterns, including tandem duplications, deletions, inversions, and translocations, served as the foundation of SV signatures [14]. The SV signatures proposed by Nik-Zainal et al. consist of 32 distinct channels, each differentiated by the type, size, and clustering status of the SVs. The consideration of the clustering status is beneficial for understanding tumor evolution, as clustered events often originate from a single, instantaneous complex

event such as chromothripsis or chromoplexy. Furthermore, a novel approach has been introduced for a more thorough assessment of SV signatures. This method involves the extraction and integration of patterns from both simple and complex rearrangements, such as duplications and triplications that are inserted distantly [33].

The framework for inferring CNV signatures in cancers was described by Steele et al. [36]. The classification schema incorporates several factors, including the number of segments for each allele at each variant loci, the heterozygosity states of the copy number segment (determining whether they are homozygous deletions, loss of heterozygosity (LOH), copy neutral LOH, or high-level amplifications), as well as the size of the segments. This classification allows for the summarization of copy number profiles using a 48-component vector, which effectively summarizes multifarious CNV states, therefore providing a better understanding of the mutational processes in a tumor.

The computational inference of SV/CNV mutational signatures involves two crucial steps. The first step is to generate mutational matrices utilizing genomic data. Following this, the matrices are analyzed to extract the mutational signatures. *SigProfilerMatrixGenerator* is a widely used bioinformatics tool that plays an instrumental role in examining the patterns of SBS, DBS and ID. It accomplishes this by converting the mutational catalogues of a collection of cancer genomes into mutational matrices. These matrices are then subjected to matrix decomposition [39]. Furthermore, the most recent version of this tool has been enhanced to facilitate the examination and matrices generation of larger mutational events, including SV and CNV signatures [40]. Several NMF-based signature extraction tools have been developed for decomposing SV and CNV signatures from matrices, such as *SigProfilerExtractor*, *pyCancerSig*, *Viola*, and *Sigminer* [26,28,41,42]. In addition to simple SV, clustering-based tools like *Starfish* have been developed to infer signatures of clustered complex genomic rearrangements using copy number and breakpoint patterns. *Starfish* has been applied to infer signatures related to biological processes by analyzing over 2000 WGS tumors. This led to the identification of three signatures associated with micronuclei- and chromatin-bridge-induced chromothripsis, as well as circular extrachromosomal DNA [43].

2.3. Applications of mutational signatures in molecular classification of cancer

Mutational signature analysis has become an integral part of the standard procedures in cancer genome analysis in both research and clinical settings. For example, although the detection of mutations in cancer predisposition genes has significantly influenced the diagnosis and optimized management of cancer patients [44], pathogenic variants in these genes are not always detectable as they can be inactivated through epigenetic mechanisms [45,46]. Moreover, assessing the pathogenicity of these germline variants can occasionally be challenging [44]. Mutational signatures have emerged as promising molecular markers, revealing previously undetected predispositions to cancer. Georgeson et al. demonstrated that the concurrent presence of two SBSs could differentiate carriers of biallelic *MUTYH* germline pathogenic variant from non-carriers in colorectal cancer (CRC) patients. This finding has shown potential for identifying biallelic carriers and classifying variants of uncertain significance [47,48]. In addition, the combination of two ID signatures was able to distinguish MMR-deficient CRCs from MMR-proficient CRCs [47]. Furthermore, Grolleman et al. identified a specific signature that could be observed in multiple malignancies carrying biallelic germline *NTHL1* mutations, highlighting the role of mutational signature analysis in characterizing tumor phenotypes in rare cancer predisposition syndromes [49]. These findings suggest that mutational signatures have the potential to serve as diagnostic tools and aid in variant classification.

The entire spectrum of somatic mutation profiles has been leveraged to accurately classify tumors with unknown origins and entities. For example, Mutation-Attention (*MuAt*), a novel deep neural network model, has demonstrated its ability to accurately predict tumor types. This model was trained by extracting informative features from the mutation data of tumor genomes from various sources including PCAWG, GEL, ICGC, and TCGA [50]. *oncoNPC*, another AI-based prediction tool, utilizes mutational signatures along with other somatic alterations and patient clinical information to jointly predict the primary origin of cancer accurately. Such predictions could potentially facilitate the clinical management of patients [51].

2.4. Application of mutational signatures in prognosis of cancer patients

In cancers characterized by complex genome alterations such as prostate cancer, it has been observed that copy number signatures exhibit a stronger correlation with clinical outcomes compared to SNV signatures [26]. CNV signatures have been utilized to predict prognosis in various cancers, such as multiple myeloma [34] and ovarian carcinoma [37]. Moreover, SV signatures are associated with distinct cancer subtypes. Adachi et al. have investigated 170 whole genomes of gastric cancer (GC) and found that non-random combinations of SV signature were associated with distinctive GC subtypes that exhibit specific driver events [52]. Yang et al. revealed the somatic SV patterns in 744 whole genome sequenced pediatric brain tumors and uncovered their role in supporting disease progression via altering cancer driver genes [53]. Understanding the mechanisms behind these alterations can not only provide unique insights into the etiology of these cancers, but also reveal potential biomarkers for patient stratification and prognosis, as well as open new therapeutic opportunities.

2.5. Application of mutational signatures in guiding treatment decision

Mutational signatures are very useful in providing information for the clinical management of cancer patients, particularly those with defects in the homologous recombination DNA repair pathway. Deficiency in homologous recombination, caused by defects in DNA repair genes, can lead to accumulated genomic instability and tumorigenesis [54]. Targeting DNA repair proteins such as poly(ADP-ribose) polymerase (PARP) in cancer cells that harbor mutations in DNA repair genes can result in synthetic lethality, providing a promising strategy for cancer therapy [55]. Furthermore, HRd has been shown to be a predictive indicator for the response to immune checkpoint inhibitors [56,57]. Several commercial tools have been developed to identify HRd. For example, the FDA-approved FoundationFocus CDx BRCA HRD, evaluates HRd by detecting alterations in *BRCA1/2* and LOH. Another HRD detector, myChoice HRD from Myriad Genetics, utilizes a combined evaluation of large-scale state transition (LST), telomeric allelic imbalance (TAI), and LOH across the genome to predict HRd status. Notably, mutational signature has been demonstrated to be a robust and independent marker of HRd [18] and is associated with response to PARP inhibitors in breast, ovarian and gastric cancers [58,59]. The identification of mutational signature is beneficial in guiding the application of PARP inhibitors in tumors that do not exhibit discernible *BRCA1/2* mutations. Given the role of mutational signatures in reflecting HRd status, various algorithms such as Signature Multivariate Analysis (*SigMA*) and *HRDetect* have been developed to determine the likelihood of HRd. *HRDetect*, which was developed based on a LASSO logistic regression model, provides a comprehensive evaluation of HRD by incorporating mutational signatures, HRD score, and deletion of microhomology. *HRDetect* has demonstrated a high predictive value in determining the response to platinum chemotherapy and PARP inhibitors. It has effectively pinpointed cases of breast cancer with HR deficiency in a phase II clinical trial [3,60–62].

Mismatch repair (MMR) plays a crucial role in maintaining genomic stability. A deficiency in MMR can result in a hypermutable state in tumors, a condition frequently observed in colorectal, gastric, and endometrial cancers [63]. Interestingly, several mutational signatures have been associated with MMRd tumors [15]. Moreover, mutational

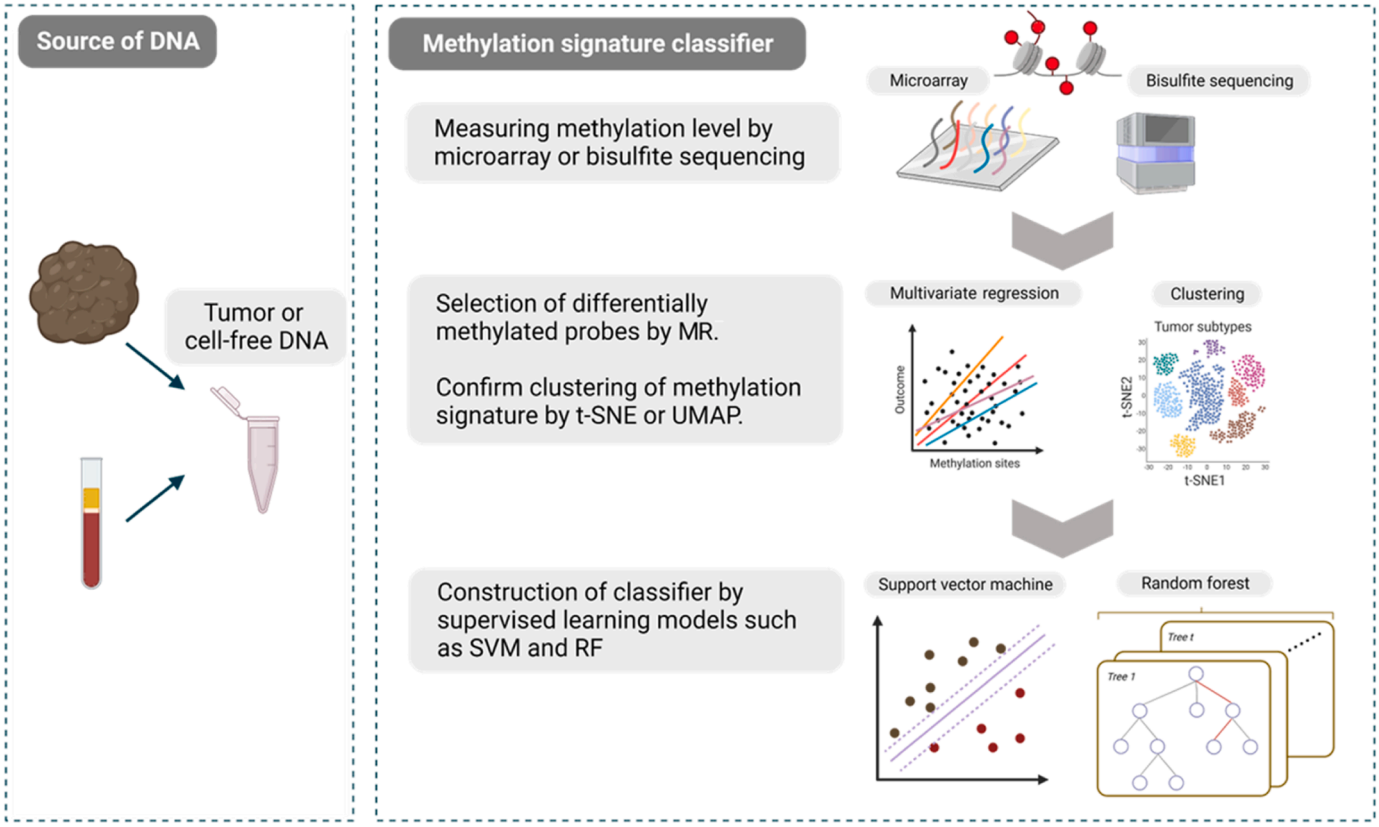

Fig. 3. Summary of general workflow of constructing the classification model from DNA methylation data. Fig. 3. The figure depicts the basic workflow of construction of DNA methylation classifier. Both tumor and cell-free DNA can be used as input. The methylation level of CpGs is measured either using targeted methylation microarray or bisulfite sequencing if genome-wide data is needed. CpGs significantly associated with the feature of interest (e.g. tumor subtype) are selected using models such as MLR. The clustering result is checked using tools like t-SNE or UMAP. Finally, the selected probes are submitted to supervised classifier-building algorithms such as SVM and RF. Abbreviation: MR, multivariate regression; SVM, support vector machine; RF, random forest.

signatures can assist in distinguishing MMRd tumors that exhibit deficiencies in different MMR proteins [19]. Notably, patients with tumors that display an MMRd signature are potential candidates for clinical treatment with immunotherapy such as pembrolizumab (a PD-1 blocking antibody) [64]. Computational tools like *MMRDetect* provide a means to analyze MMRd in a genome by examining the spectrum of mutational signatures. This approach complements the current method of directly sequencing causal genes such as *MSH2*, *MSH6*, *PMS2*, and *MLH1* [65].

Genetic alterations are extensively studied for their potential as indicators of drug sensitivity. Recent findings suggest that mutational signatures provide a more accurate prediction of a cell line's response to drugs, compared to solely examining gene mutations in oncogenic driver genes. Levatić et al. have identified several mutational signatures associated with drug activity across various cancer cell lines. For example, several MMRd signatures are associated with sensitivity towards AKT serine/threonine kinase inhibitors. Moreover, signatures associated with previous chemotherapy exposure tend to correlate with resistance to future drug treatments [66].

3. DNA methylation signature in cancer

DNA methylation is the most common modification of the human genome which is found primarily at the cytosine of CpG dinucleotides. This process is under strict regulation by the methyltransferases *DNMT3A, DNMT3B, DNMT3L* and methylcytosine dioxygenases TET [67]. The distribution of methylated CpG along the human genome is not uniform. Most of the human genome is CpG-poor and methylated in differentiated somatic cells. However, several thousands of short interspersed CpG-rich sequences, known as CpG islands, locate within gene promoters and are hypomethylated with unique tissue-specific patterns. Early functional studies demonstrated that DNA methylation induced transcriptional repression [68,69]. In cancer, DNA methylation is heavily studied for its roles in regulating gene transcription and genome instability [70]. Many studies have demonstrated that cancer cells have a significantly different DNA methylation profile compared to normal cells. In general, cancer cells display global hypomethylation and local hypermethylation especially around tumor suppressor genes [71]. Hypomethylation is shown to promote genomic instability, activation of transposable elements as well as aberrant gene expressions [70]. Being itself one of the hallmarks of cancer, DNA methylation and epigenetic reprogramming have essentially contributed to all other cancer hallmarks including resistance to cell death, dysregulation of cellular metabolism and escaping from immune destruction [72]. The methylation landscape of cancer cells is often tissue- and subtype-specific. This specificity is shown to arise potentially from the effect of somatic mutations [73,74] and interference from the diverse tumor microenvironment such as hypoxia, stomal cells and infiltrating immune cells [71,75].

3.1. Detecting methylation signature

In recent years, next-generation sequencing (NGS)-based or array-based technologies are the most popular methods to study genome-wide DNA methylation (Fig. 3). Whole genome bisulfite sequencing enables genome-wide investigation of DNA methylation profiles at a single base resolution while high-throughput methylation array allows robust DNA methylation analysis in a cost-effective manner [76,77].

Although methylation array can have variable coverage over whole genome based on the probe design, it is currently the most frequently used technology for DNA methylation studies, especially in clinical settings [78,79].

After obtaining the DNA methylation profile of a cohort of samples, different computational processes can be applied to establish a disease-specific episignature. The first step is usually selection of CpG sites that are differentially methylated in patients compared to healthy controls. The selected sites are then used to train and construct classifier using statistical and machine learning methods to distinguish patients from controls [80,81]. Methylation CpG site selection is usually achieved by statistical methods such as multivariable linear regression modeling and sometimes additional checking of potential non-normal distribution of methylation signal such as non-parametric Mann-Whitney U test may also be needed [82,83]. During this process, hundred-thousands of CpG sites across the genome are reduced to few hundreds of significantly differentially methylated sites. This avoids overfitting of the classifier model and reduces the complexity of the model [80]. CpG sites correlating significantly to other CpGs and CpGs with small effect size can be removed for further enrichment [84]. According to some studies, including estimated blood cell type distribution into the regression model helps refining it [85,86].

The episignature generated includes assessed and visualized using different methods including hierarchical clustering analysis, t-distributed stochastic neighbor embedding (t-SNE) and uniform manifold approximation and projection (UMAP) to confirm its efficiency [87,88]. The episignature(s) will then be used for construction of classification algorithm. Examples of common algorithms are support vector machine (SVM), elastic net multinomial logistic regression (ELNET) and random forest (RF). Benchmarking studies have been performed to compare these algorithms for their performance in cancer DNA methylation data and all of the three in general had satisfactory results [8,89]. For example, Maros et al. showed that while ELNET performs the best as a stand-alone algorithm, SVM and RF can achieve similar results after model-updating calibration using ridge-penalized multinomial logistic regression [89].

3.2. Applications of methylation signatures in molecular classification of cancer

Owing to the tissue- and subtype-specific nature of DNA methylation in cancer, much effort has been made to utilize it in molecular classification and cancer diagnosis. In medulloblastoma, the disease was originally classified into 4 subgroups by its transcriptional profile [90]. While group 1 and group 2 were characterized by activation of SHH and WNT pathways respectively, groups 3 and 4 were later shown to be distinguishable by their methylation signatures [91]. This eventually led to the inclusion of group 3 and 4 medulloblastomas in the 2021 WHO classification of tumors of the central nervous system (CNS) based solely on molecular features [92]. The first few studies used traditional statistical methods for classification such as unsupervised hierarchical clustering. Capper et al. extended the application of such methylation-based classifier to all subtypes of CNS tumors using a large cohort ($n = 2801$) and the random forest approach as mentioned in Section 2.1. The classifier is available at (https://www.molecularneuropathology.org). It was benchmarked with histopathological evaluation and high concordance (838/1104, 76%) was observed. Most importantly, it helped the diagnosis of CNS tumors in patients with no clear histological-definable subgroup and provided a new diagnosis in a significant number of cases [81]. This approach is also widely applied on many types of cancers including sarcoma, colorectal cancer, and breast cancer [93–95]. In some of these classifiers, they also aimed at stratification of patients based on prognosis, risk of relapse, and treatment response [96–98].

3.3. Application of methylation signatures in early detection of cancer

The application of episignature in early non-invasive screening of cancer is another active area of research. Some of the earliest studies on colorectal cancer discovered that aberrant methylation around oncogenes and tumor suppressors occurs at early stage of carcinogenesis [99,100]. Imperiale et al. subsequently validated the methylation biomarkers in a large-scale clinical trial showing promising sensitivity and specificity [101]. DNA from colorectal cancer cells can be readily examined by using fecal DNA samples. However, in most other cancers, only cell-free DNA (cfDNA) from peripheral blood is available. Moreover, in some cancers, the specificity requirement of methylation biomarkers is not only limited to distinguishing between early cancer patients and healthy individuals. For example, in hepatocellular carcinoma which often originates in patients with chronic cirrhosis, methylation biomarkers need to differentiate early HCC patients from cirrhotic patients [102]. In recent years, a number of in-vitro diagnostic (IVD) tests based on DNA methylation analysis of cfDNA have been developed. The majority of them target a cancer-specific panel of oncogenes and tumor suppressors [103–105].

With the help of whole genome methylation sequencing and more advanced computational methodologies, recent studies also focused on building more comprehensive methylation signatures to develop a pan-cancer screening assay [106–109]. GRAIL, a biotechnology company focusing on early cancer screening, in collaboration with the Mayo Clinic, designed the Circulating Cell-free Genome Atlas (CCGA; NCT02889978) study to combine genome-wide cfDNA sequencing with machine learning to develop pan-cancer classifiers with high specificity. The study aimed at recruiting approximately 10,500 cancer participants and 4500 non-cancer controls [110]. Among all the molecular features tested, whole-genome methylation outperformed other genetic markers such as small somatic variant panels and whole-genome somatic copy number alterations [110]. Two ML modules were trained – one to determine cancer/non-cancer status and the other to predict tissue of origin. The locked classifier currently covers 103,456 distinct genomic regions (17.2 Mb) and 1116,720 CpGs features with a specificity of 99.5% and an overall sensitivity of 51.5% [108]. Recent validation study by Nicholson et al. also revealed similar findings and demonstrated that the sensitivity correlates positively with tumor stage and varies among different cancer types with an overall sensitivity of around 20% in stage I patients [111].

3.4. Application of methylation signatures in prognosis of cancer patients

Methylation signature of tumors can be used to reveal the molecular heterogeneity of the tumors in terms of their proliferative potential, resistance to apoptosis and invasive ability. Therefore, methylation signature is demonstrated to serve as a prognostic marker for cancer patients. Bladder EpiCheck, a methylation test measuring 15 methylation probes which was originally designed to examine the presence of bladder cancer using urine samples, can be used to monitor cancer recurrence after surgery [112]. This test, similar to those mentioned in the above sections, had a high specificity and negative prediction value, which made it a good choice for surveillance to exclude recurrent disease [113]. In a few other tests such as the Colvera assay which was developed for monitoring relapse in colorectal cancer patients, only very few methylation probes were measured [114,115]. For monitoring cancer recurrence, methylation signature can sometimes just serve as an indicator of presence of tumor cells based on cancer-specific differentially methylated CpGs irrespective of their biological functions.

Besides monitoring for recurrence, methylation signature has also been shown to predict survival of patients. For example, a DNA methylation signature consisting of ten differentially expressed genes in gastric cancer predicts both overall survival and relapse independent of TNM stage [98]. In another study using a cohort of 1538 breast cancer patients, Batra et al. developed a semi-supervised computational

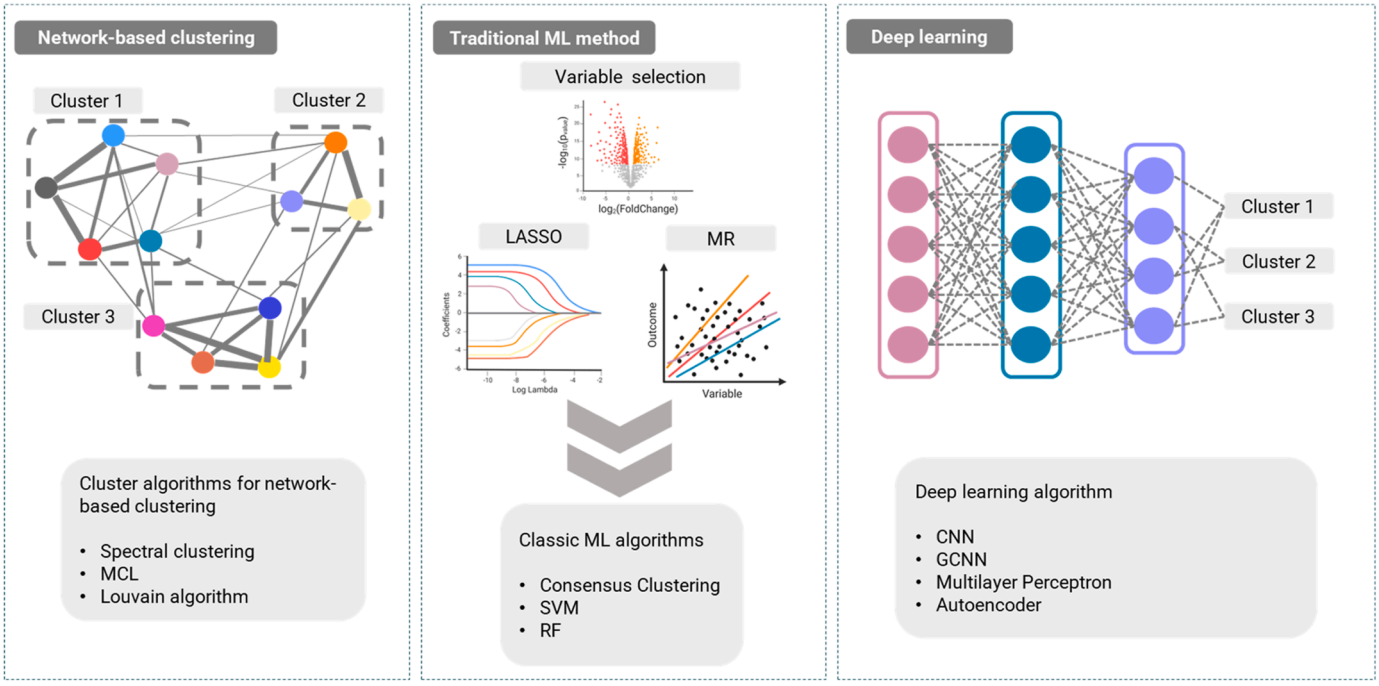

Fig. 4. Examples of clustering algorithms in analyzing transcriptome signature. Fig. 4. The figure shows examples of unsupervised clustering algorithms used in identifying molecular subtypes of cancers using transcriptomic data. Network-based algorithms such as spectral clustering, MCL and Louvain identifies clusters without making prior assumptions about the nature of clusters in the data. Traditional ML methods start with selection of subtype associated DEGs using regression models, followed by submitting them into ML algorithms like consensus clustering, SVM and RF. Deep learning models gain increased interests in recent years. Whole-transcriptome data can be used directly in training models like CNN and GCNN without the need for prior filtering of DEGs. Abbreviation: CNN, convolutional neural network; DEG, differentially expressed gene; GCNN, graph convolutional neural networks; LASSO, least absolute shrinkage and selection operator; MCL, Markov cluster algorithm; ML, machine learning; MR, multivariate regression; SVM, support vector machine; RF, random forest.

strategy (Methylayer) which integrates different biological features including gene expression, DNA mutational spectrum, and clinical information for computational peeling of confounders to model layered tumor methylation signatures. They identified that factors of the tumor microenvironment (TME) such as infiltrating immune cells and cancer-associated fibroblast strongly interfered with a subgroup of methylation sites [116]. Interestingly, they also introduced the idea of epigenomic instability which is made up by differentially hypermethylated and hypomethylated CpGs (which they termed methylation gain and methylation loss layers respectively) and showed that they predicted overall survival oppositely in ER negative breast cancer patients [117]. On the other hand, they also showed that there existed a third group of differentially methylated CpG sites whose level did not correlate with tumor stage and found that this third group did not have any prognostic value. This study demonstrated that methylation signatures can be further refined by including other potential confounders like factors of the TME. Similar models for predicting prognosis have also been developed for other cancers including hepatic, pancreatic, colorectal cancers, and meningiomas [118–121]. Furthermore, methylation signatures were able to detect patients with highly metastatic tumors. Wu et al. and Chen et al. constructed methylation models to classify patients with potential lymph node metastasis in gastric cancer [122,123]. Other studies also revealed different methylation signatures associated with distant metastasis in cancers including the prostate, colon, and lung [124–127].

3.5. *Application of methylation signatures in guiding treatment decision*

Given the effect of genome-wide methylation patterns in regulating gene expression in cancers, many groups tried to use methylation signature in tumor cells to predict drug responses. By just identifying the tissue of origin for carcinoma with unknown primary (CUP), it has been shown that patients who received tumor type-specific chemotherapy had an improved overall survival compared to those receiving empiric therapy [128,129]. Furthermore, with the initial success of *MGMT* promoter methylation analysis in guiding temozolomide treatment in gliomas [130], methylation signature analysis can now be applied to study the effect of any anti-cancer agents in a pan-cancer manner. Iorio et al. studied different molecular signatures in 1001 human cancer cell lines across 29 tissues and demonstrated that methylation signals in combination with genomic data were the best tissue-specific predictor of drug responses in 120 of 319 tested anti-cancer agents [131]. Moreover, methylation signatures were demonstrated to be effective predictors of responses to both chemotherapies and targeted therapies in different types of cancers [132–134]. In recent years, as immunotherapies such as checkpoint inhibitors were proven to be superior to conventional chemotherapies in many cancers, many studies have also focused on investigating if methylation signature can serve as a predictor for immunotherapies. Duruisseaux et al. developed the EPIMMUNE signature, a methylation signature extracted from non-small-cell lung cancer DNA samples which can predict response to anti-PD-1 treatment [135]. The authors suggested that DNA samples from bulk tumor samples had the advantage of being able to reflect intrinsic (cancer cell) and extrinsic (microenvironment and infiltrating immune cell) factors of the tumor which both govern the resistance to immunotherapy. Subsequent studies further extended the use of methylation signatures in predicting response to immunotherapy in melanoma [96] and even in a pan-cancer manner although with reduced specificity [136].

4. Transcriptome signature in cancer

The transcriptome reflects the physiological state of the cell. Transcriptome-wide study of gene expression began in the 2000's starting with studying those of established cell lines and healthy subjects [137,138]. Much research on characterizing the transcriptome cancer tumors and cell lines emerged at that time [139–141] and in the early

2010's when NGS technologies were adopted, for example, in renal adenocarcinoma [142,143], lung cancers [144] and pancreatic cancers [145]. These earlier studies focused on discovering alterations of the transcriptome of the cancer of a given origin, such as alternative splicing events in lung cancer [144], the *ALK-PTPN3* gene fusion in non-small cell lung cancer (NSCLC) [146] and the *RB1-ITM2B* fusion in melanoma [147].

4.1. Applications of transcriptome signatures in molecular classification of cancer

Transcriptome has also been widely used for discovering differentially expressed genes (DEGs) in cancers. It was initially proposed that identification of key sets of DEGs (transcriptome signature) can help in diagnosing patients with CUP [148]. The complexity of the structure of transcriptome is contributed by the fact that gene expression is often regulated by a network of co-related genes which interact and form clusters and networks. Different statistical methods were developed to assist in computing the clusters of DEGs and identifying the characteristic transcriptome signatures, which include vector algebra-based algorithm [149], singular value decomposition [150] and principal component analysis [151]. The field then gradually switched to the use of machine learning approaches such as SVM and deep neural network given their superior performance in decoding the hidden patterns in transcriptome (Fig. 4). [152,153]. These new classifiers can in general achieve an overall accuracy above 90% and support classification of more than 30 types of tumors from different origins [154,155].

Transcriptome signatures have also been investigated as a potential tool to help in molecularly subtyping within the same type of cancer. Guinney et al. applied an unsupervised network-based cluster algorithm to identify network substructures which correspond to different molecular subtypes of colorectal cancer [156]. They showed that colorectal cancer can be molecularly grouped into four consensus subtypes with distinct intrinsic and microenvironmental features. In addition, Joanito et al. showed that by further isolating the transcriptome signatures of the epithelial cells populating the tumor microenvironment using single-cell sequencing, they managed to refine the classification of colorectal cancer subtypes based on results of bulk tumor transcriptomes [157]. Following a similar methodology, consensus molecular subtypes were identified in muscle-invasive bladder cancer and non-small-cell lung cancer [158,159]. In some studies, this clustering process is iterated using resampling-based algorithms like consensus clustering [160] which aim at reaching a stable consensus model representing results over multiple runs of clustering [161,162].

4.2. Application of transcriptome signatures in early detection of cancer

There is research on circulating tumor RNA in readily accessible saliva or epithelial cells for early detection of cancers. The transcriptome signature in saliva has been proposed to be utilized in early detection of oral squamous cell cancer [163,164], with a clinical trial conducted in 2015 with a receiver operating characteristic AUC of over 0.85 [165]. For lung cancer, Whitney et al. developed a classifier from the transcriptome of bronchial epithelial cells in a population of current or former smokers using multi-step logistic regression and obtained an AUC of 0.80 [166]. Silvestri et al. applied this classifier to a separate cohort and obtained AUC of 0.78 [167]. It improved the diagnostic yield over bronchoscopy alone while having a negative predictive value of 91%. This population contained both early-stage and late-stage primary lung cancer, and all stages were detected using the classifier. Recently, Mazzone et al. explored the use of nasal epithelial cells for non-invasive screening but positive predictive value was modest [168].

4.3. Application of transcriptome signatures in prognosis of cancer patients

Machine learning is yielding new transcriptome signatures independent of existing signatures. In the case of lung adenocarcinoma, Xu et al. recently found that the combination of random survival forest (RSF) and generalized boosted regression modeling algorithms yielded an optimal model of 52 overall-survival-associated genes with mean C-index of 0.692 in 11 cohorts [169] that stratifies patients by OS and is independent of 108 published signatures and greater precision than clinical features such as age and cancer stage. In addition, they found the high-risk group was sensitive to alisterib, while the low-risk group was sensitive to RITA, which may guide treatment decision. For acute myeloid leukemia (AML), discriminative models are needed for improving the current European Leukemia Net classification system. Selected expression-based models have been assessed by Wang et al. [170]. Recently, narrowing down the genes based on the cancer's pathogenesis has proven successful in generating prognostic models. Tao et al. developed a prognostic model from 39 genes and 8 lncRNAs involved in ferroptosis in pediatric AML [171]. The risk score stratified patients with high and low overall survival with an AUC of 0.70 for 1-year and 5-year survival. In the case of colorectal cancer, Samadi et al. found novel prognostic biomarkers from analysis of mRNA, lncRNA and miRNA data using Robust Rank Aggregation (RRA) and WGCNA [172]. A recent review by Tran et al. includes more examples of ML-derived prognostic signatures [173].

4.4. Application of transcriptome signatures in guiding treatment decision

The transcriptomic characteristics of malignancies have proven useful for clinical risk stratification and guiding treatment decisions especially in leukemia patients [174–176]. A study by Docking et al. demonstrated that an expression-based prognostic score can provide more accurate risk stratification for AML patients which can help direct choice of treatment. Furthermore, they showed that transcriptome-based testing can enhance therapy selection by identifying a subset of patients with dysregulated integrin signaling [176]. Importantly, transcriptomic profiling was demonstrated to be valuable in immuno-oncology. RNA-sequencing data contributed to the development of personalized cancer vaccines by characterizing human leukocyte antigen allotype and clonal expanded antitumor T cells [177,178]. On the other hand, transcriptome signature of adaptive immunity has been demonstrated to be used as predictive markers of response to immune checkpoint blockade therapy [179].

5. Proteomic and metabolomic signatures in cancer

Proteomics and metabolomics, as effectors of genomics and transcriptomics, have emerged as important fields in cancer mechanistic research to identify clinically applicable biomarker signatures. In the last decade, advances in mass spectrometry (MS) have enabled precise profiling of cancer proteomes and metabolomics at cell, tissue, and biofluids. In light of the successes with MS-based technologies in signature discovery [180], more readily accessible methods have been developed to facilitate the discovery. In recent years, affinity-based high throughput proteomic profiling using antibodies or aptamers, which are mostly applicable to liquid biopsies, has revolutionized proteomic cancer research [181].

5.1. Applications of proteomic and metabolomic signatures in molecular classification of cancer

Together with the improvements in machine learning algorithms, recent studies have demonstrated that proteomic signatures can be used for determination of tissue of origin in patients with CUP [182,183]. However, it is less well-validated compared to methylation signatures

and still requires further studies to increase its sensitivity and specificity.

Significant efforts have been made to incorporate proteomic signature into other multi-omics data to refine the molecular subtyping of cancers. Two international networks, The Clinical Proteomic Tumor Analysis Consortium (CPTAC) [184] and The International Cancer Proteogenome Consortium (ICPC), have been established to promote collaboration between proteomic groups and other genomic groups. Part of their work included standardization of bioinformatic analysis protocols of proteomic data [185]. In the early studies, statistical clustering methods such as model-based clustering, consensus clustering and machine-learning classifiers were applied to discover molecular subtypes based on proteome results alone [186,187]. In subsequent CPTAC and ICPC studies, multi-omic clustering using NMF became the most popular tool for integrative analysis of proteomic signature together with other multi-omic signatures in subtype classification [188–191]. While the majority of subtype classifications were concordant between integrative proteomic signature and other omics signatures, a subset of tumors was reclassified by proteomic signature. Since proteomic signature measures the final functional protein quantities after post-transcriptional and post-translational regulations, the reclassification revealed intra-subtype heterogeneity. Moreover, it also provided valuable information in identifying new druggable protein targets especially with the phosphoproteome data [191,192].

Several metabolic alterations occur in cancer cells, serving as a new hallmark, in which metabolites serve as substrates for energy generation and biomass formation as well as regulators of transcriptomes and proteomes that affect the tumor micro and macroenvironment [193]. Efforts have also been made to develop cancer subtype classification based on metabolomics signatures using various ML methods and strategies. For instance, a study by Gal et al. stratified breast cancer cases based on tumor-tissue metabolome profiles and identified three distinct subtype clusters using specific metabolite profiles and associated with different tumor stages and prognosis survival stages using the k-sparse ML method among the five tested [194]. As another example, Machlin et al. utilized urine metabolic profiles and developed a prediction model for bladder cancer that distinguished the disease from controls and further stratified it by grades [195]. They subtracted the potential cofounder effects by gender and age, and built the model based on three metabolites, having an AUC of 0.956 for bladder cancer risk prediction. Tan et al. used a binary logistic regression model with a stepwise optimization algorithm to identify the three most effective differential metabolites for constructing a diagnostic model for bladder cancer with an AUC of 0.961 [196]. Metabolomic biomarker signatures have become increasingly important for classifying cancer subtypes and predicting clinical outcomes [197–200], with novel methods continually emerging.

5.2. Application of proteomic and metabolomic signatures in early detection of cancer

The earliest studies in proteomic signature for cancer diagnosis gained focus in the early 2000 s when Petricoin et al. and Adam et al. developed proteomic classifiers for diagnosis of ovarian cancer and prostate cancer respectively [201–204]. They demonstrated pattern analysis using machine-learning technique (self-organizing map/Kohonen network) [205] can serve as an alternative method to single biomarker discovery in high dimensional proteomic studies [201]. Since then different classifier training algorithms have been used for identification of diagnostic protein signatures while no single type of technique consistently outperformed the others. It is common that multiple methods were used in a single study and the investigators eventually selected the best-performing model [206,207].

Recently, large-scale biomedical databases, like UK-biobank and FinnGen, provided a wealth of resources for discovery of quantitative trait loci across various omics and expression levels, useful in detecting risk of disease, such as cancer, at an early stage [208,209]. Analysis of large-scale population-based proteome or metabolomic profiles in conjunction with genome-wide disease-association data through the two-sample Mendelian randomization approach (MR) allows mapping of disease-causative protein- or metabolite-QTLs that could be used as risk predictors [210]. The MR approach is increasingly being utilized to discover biomarkers and QTLs for a wide variety of diseases [211–213], including cancers, because it is capable of establishing causal relationships between expression profiles and diseases while overcoming the limitations of residual confounding and reverse causality. For example, Mälarstig et al. first identified 812 cis-pQTLs of 737 proteins for seven breast-cancer-related clinical characteristics (e.g. age, alcohol consumption and number of births) based on plasma proteomes of 2929 proteins in 598 women using linear regression model [214]. By applying the cis-pQTLs as genetic instrument and the 730 proteins as exposures, they performed both Wald-ratio and Inverse-variance weighted (IVW) MR analyses on breast cancer disease-association datasets from three independent sources and identified five proteins that are etiologically relevant for breast cancer development. In another example, Sun et al. used a similar approach to identify 13 circulating proteins for CRC risks based on GWAS meta-analysis datasets from literatures and from UK-biobank and FinGen [215]. They also examined the pleiotropy of the protein effects on CRC risk using MR-Egger regression method [216] which was one of the methods developed to deal with the contamination of invalid instrumental variables [217–219]. Similarly, Feng et al. carried out a two-sample MR analysis based on serum metabolomic profiles and GWAS data of multiple cancers and identified key biomarkers for each cancer [220]. The cancer-relevant QTLs and plasma biomarkers provide a powerful tool for identifying individuals at high risk at an early stage of cancer.

5.3. Application of proteome and metabolomic signatures in prognosis of cancer patients

In addition to discovering novel molecular subtypes of cancers, the works of CPTAC and ICPC also focused on investigating whether proteomic signature may offer extra prognostic value to existing multi-omic signatures. For example, Krug et al. found that the integrative proteomic signature identified a subgroup of PAM50 luminal A breast cancer patients to be luminal B-like (which they donated as NMF LumB-I) which was associated with poor overall survival than NMF LumA-I [190]. It should be noted that integrative analysis of proteomic signature may not always produce similar results as stand-alone analysis of proteomic data. This can be particularly highlighted by the study by Asleh et al. in which they further subclassified the triple-negative breast cancer (TNBC) patients into 4 biological subgroups with differential recurrence-free survival solely by consensus clustering using proteomic data [221]. This heterogeneity in TNBC was not revealed by the NMF analysis used by Krug et al. Similarly in lung adenocarcinoma, Soltis et al. developed a proteomic signature to predict overall survival and metastasis-free survival which differs from the result of somatic genome signature subtyping [222].

Longitudinal metabolomics serves as a potential approach to identify signatures for monitoring cancer progression, relapse and remission using regression models and ML models [223,224]. Among them are TCA cycle intermediates and RNA degradation products for CRC [225], and carnitine and acetylcarnitine for multiple myeloma [226]. Since the patients are likely receiving therapy during monitoring, understanding the therapeutic drug metabolism is important to distinguish between treatment-induced metabolic changes and cancer-relapse-induced metabolic changes.

5.4. Application of proteomic and metabolomic signatures in guiding treatment decision

Proteome and phosphoproteome signatures are capable of directly measuring the degree of activation of target oncogenic pathways, making them sensitive predictors of response to targeted therapies

Table 1
Examples of classification methods and their performance in different omic signatures.

Classifier method	Type of signature	Clinical application	Performance	Reference
Logistic regression	Methylation signature	Early detection	AUC = 91%, sensitivity = 98%, specificity > 99% for ovarian cancer[106]; Sensitivity = 44.2%, specificity = 99.8% for stage I–III cancers[107]	[106,107]
	Methylation signature	Predict treatment response	Overall survival HR = 0.080, p = 0.0012	[135]
	Transcriptome signature	Determination of tissue of origin	Accuracy = 0.911, sensitivity = 0.800	[155]
	Transcriptome signature	Early detection, subtyping	AUC = 0.78, sensitivity = 0.93, specificity = 0.57[166]; Accuracy = 0.92[261]	[166,261]
	Transcriptome signature	Prognosis	Overall survival HR = 4.11, p < 0.001	[262]
	Metabolic signature	Early detection	AUC = 0.838, sensitivity = 0.807, specificity = 0.818	[263]
Graph-based clustering	Methylation signature	Subtype classification	5 subgroups identified	[264]
	Transcriptome signature	Subtype classification	2 subgroups identified[157]; 6 subgroups identified[158]	[157,158]
Kernel logistic regression	Methylation signature	Early detection	Sensitivity = 0.34, specificity = 0.98	[110]
LASSO	Methylation signature	Predict treatment response	Risk groups stratification, Kaplan–Meier plot p = 0.0034	[134]
	Transcriptome signature	Predict treatment response	Risk groups stratification, HR = 5.46, p = 4.57e− 13	[176]
	Proteomic signature	Prognosis	Risk group stratification C-index = 0.885, p < 0.0001; AUC = 0.95	[265]
Support vector machine	Methylation signature	Predict treatment response	F1 score = 0.4255, AUC = 0.6742	[136]
	Transcriptome signature	Subtype classification	Accuracy = 0.78	[152]
	Proteomic signature	Response to Immunotherapy	AUC = 0.77	[237]
Random forest	Methylation signature	Subtype classification	91 subgroups identified, estimated error rate = 4.28%[81]; 65 subgroups identified, estimated error rate = 0.65%[93]	[81,93]
	Transcriptome signature	Subtype classification	Accuracy = 0.854, sensitivity = 0.859, specificity = 0.94	[156]
	Transcriptome signature	Prognosis	C-index = 0.692	[169]
NMF clustering	SNV, CNV and SV mutational signature	Extracting *de novo* signatures	NA	[27,28,33,36]
	Proteomic signature	Subtyping	3 subgroups identified[189]; 4 subgroups identified[191]	[189,191]
Unsupervised k-means clustering	Methylation signature	Prognosis	Overall survival HR = 1.92, p = 0.01	[102]
	Transcriptome signature	Subtype classification	4 subgroups identified	[161]
	Proteomic signature	Prognosis	Overall survival, cluster 2 HR = 2.68, p = 0.001	[221]
Agglomerative hierarchical clustering	Methylation signature	Prognosis	Risk group stratification OR = 15.45, p ≤ 0.05	[121]
	Transcriptome signature	Subtype classification	8 subgroups identified	[162]
	Metabolic signature	Predict treatment response	Overall survival, S2 subtype HR = 4.187, p = 0.002	[239]
Neural network-based clustering	CNV and SV mutational signature	Extracting *de novo* signatures	6 complex genomic rearrangement signatures identified	[43]
	Mutational signature	Predict treatment response	AUC = 0.98	[266]
	Transcriptome signature	Subtype classification	Accuracy > 80% in breast cancer cohort	[94]
	Transcriptome signature	Diagnosis	Accuracy = 97%, mean F1 score = 0.92	[267]
	Transcriptome signature	Subtype classification	Sensitivity = 0.9733, specificity = 0.9737, F1 score = 0.9733[153]; Accuracy = 0.766, AUC = 0.852[259]	[153,259]
	Proteomic signature	Early detection	Sensitivity = 1.00, specificity = 0.95	[204]

[227–230]. Since liquid biopsies, such as blood and urine, can be accessed with minimal invasiveness, they are ideal sampling types for monitoring disease progression monitoring and managing treatment. At present, the most widely validated serum proteomic signature predictor is the VeriStrat test for predicting the benefit of EGFR inhibitor erlotinib treatment for NSCLC patients [231–233]. It was first developed based on serum proteome profiles using a straightforward k-nearest neighbor (KNN) algorithm [231]. Later, it was demonstrated to be significantly associated with proinflammatory NSCLC [234]. Chae et al. and Rich et al. therefore proposed that VeriStrat could be a potential signature marker for predicting the response to immunotherapy in NSCLC patients [235,236] and demonstrated it to be an independent survival predictor in addition to PD-L1 tumor expression, suggesting proteomic signature could be a superior predictor than standard biomarkers for immunotherapies [236]. The VeriStrat for NSCLC subtype classification and treatment decision suggestion demonstrated the power of proteome signature analysis in predicting treatment responses. Recently more proteomic signatures specific to predict response to immunotherapy were developed in different cancers [237,238].

Spatial omics analysis enriches our understanding of the intratumoral heterogeneity and its interactions with the adjacent stromal tissues, which can be beneficial for therapeutic decisions. A recent study by Wang et al. detected differential spatial metabolomic signatures in lung squamous cell carcinoma between the stroma tissues and tumor tissues and found that the stromal metabolomic profiles associated with chemotherapy response, providing insight for immunotherapies [239].

6. Overview of each signature and future directions in multiomics

Previous studies have focused on driver mutations in the coding region that alter protein function and drive tumor cell survival and growth. A subset of these driver mutations has been successfully targeted therapeutically. For instance, gefitinib, erlotinib, and afatinib have been used to treat lung cancer patients carrying *EGFR* mutations, and larotrectinib and entrectinib have been approved for the treatment of solid tumors carrying *NTRK* fusions. Considering that driver mutations are not always found in every cancer genome and the majority of cancer variants are passenger mutations located in non-coding regions, the emergence and evolution of mutational signatures have bridged the gap in understanding the whole genome and pinpointing therapeutic targets. Mutational signatures have demonstrated significant potential in clinical settings, both in terms of tumor classification and as prognostic indicators. Signatures associated with known etiologies, such as HRD and defective MMR, have shown significant potential in clinical applications, offering opportunities for the use of PARP inhibitors and

immunotherapies. These provide a strong justification for considering mutational signatures alongside driver mutations when interpreting cancer genomes. Although the emergence of large amounts of genome-wide sequencing data has enabled the discovery of common signatures shared by different types of cancer as well as cancer-type specific mutational signatures, the etiology of many of these signatures remains elusive, future works that utilize cell lines and model organisms, are crucial for validating the functional outcomes and facilitating the interpretation of these mutational signatures.

Given the importance of DNA methylation in early detection, the prediction of cancer progression and metastasis, clinical outcomes, and response to therapy, it plays a vital role in guiding cancer clinical management. A variety of methods have been developed to provide a comprehensive view of the DNA methylation landscape. These include genome-wide bisulfite sequencing, methylated DNA immunoprecipitation (MeDIP), and Illumina 450 K arrays, which measure the methylation state of well-characterized CG sites distributed across the genome. DNA methylation signatures have been shown to be effective diagnostic and prognostic markers (Table 1). For example, machine learning-based approaches have enabled the classification of central nervous system tumors and the differentiation of primary lung squamous cell carcinomas from head and neck metastases, thereby informing therapeutic decisions [240]. Moreover, DNA methylation patterns derived from FFPE from glioblastoma patients can be used to predict patient survival [241]. Besides DNA methylation signatures, we have summarized the application of transcriptome signatures in diagnosis, prognosis, and guiding treatment decision for cancer patients. However, transcriptome signatures from heterogeneous tissue are known for their limitation in accurately reflecting the cell type-specific characteristics of tumor tissue. Moreover, only a limited number of methylation and transcriptome tests have received FDA approval for clinical use. The critical challenges include overcoming cell-type heterogeneity, developing reliable and standardized methods for selecting methylation/transcriptome features to predict clinical outcomes and therapeutic responses, and assessing the clinical utility and reliability of the identified signatures.

DNA mutational signatures, methylation signatures, and transcriptome signatures provide useful information from different perspectives. Multi-omics that integrate data across multiple layers have the potential to further enhance the performance of existing models. For example, transcriptome sequencing by RNA-seq allows for the detection of altered expression, splicing and gene fusion events. The incorporation of RNA-seq with genetic testing (target panels, WES or WGS) has been shown to enhance the detection, classification, and validation of disease-causing variants, and provide treatment alternatives in cancer patients [242–246]. In addition, the concurrent profiling of the DNA methylome with the whole genome and transcriptome has been shown to substantially improve the detection efficacy of clinically actionable variants in pediatric cancer patients [247]. The multi-omics profiling encompasses transcriptome and DNA methylome, showing superior predictive performance in classifying cancer subtypes among breast cancer, glioblastoma, and ovarian cancer, compared to using single omics data [248]. Furthermore, multi-omics profiling that includes multiple layers of data, such as WGS, RNA-seq, Hi-C and ATAC-seq has been shown to identify changes in the three-dimensional organization of the genome, chromosome accessibility, and gene expression. This comprehensive approach provides valuable insights into the regulatory mechanisms of aberrant gene expression and disease progression [249,250].

The field of multi-omics has experienced significant advancements, particularly in the transition from bulk analysis to single-cell analysis. Signature detection based on bulk sequencing largely depends on the cellular composition of the sample and represents an average molecular signal. To tackle the challenges of tumor heterogeneity, single-cell techniques have emerged, facilitating personalized treatment based on the specific heterogeneity of the tumor. Single-cell RNA-seq has proven successful in dissecting tumor heterogeneity and microenvironment with unparalleled resolution in both solid tumors and hematological malignancies such as leukemia and lymphoma [251]. For example, scRNA-seq has been utilized to identify drug-tolerant cell populations in NSCLC tumors, as well as quiescent stem-like cells that contribute to chemoresistance and poor outcomes in AML. These findings have provided targets for eliminating these cell populations to overcome therapy resistance [252,253]. Furthermore, it is currently under active investigation in clinical trials in hematology and oncology, with the goal of discovering biomarkers, enhancing diagnostics, and refining disease subclassification to improve patient care [254]. On the other hand, single-cell DNA sequencing (scDNA-seq) has demonstrated significant potential in early diagnosis and disease monitoring. This is exemplified by its ability to detect measurable residual disease with a high sensitivity of approximately 0.01%. Moreover, it simultaneously provides clone-specific immunophenotypic data in cases of acute leukemia [255,256]. The integration of single cell RNA/DNA-seq with other single-cell assays such as scTCR/BCR-seq, scATAC-seq and scHi-C, allows for the simultaneous study of not only the genotypic and phenotypic characteristics of individual cells, but also the underlying regulatory mechanisms in cancer.

Multi-omics analysis of cancer tissues provides valuable insights into mechanistic understanding of cancer, however such invasive samples are less suited for regular disease monitoring. Cancer-specific changes in cellular metabolism and proteome, which manifest in the form of secreted signatures and present in the circulating blood, offer an ideal and less invasive means for clinically monitoring the disease progression [223]. An important consideration during data interpretation is the metabolic and proteomic profiles of cellular uptake and excretion from all bodily processes, not only from the cancerous process. Longitudinal monitoring would provide dynamic omics changes in trends and greatly enhance prognostic monitoring. While the sequencing technologies for genome and transcriptome enable nearly complete coverage, proteome and metabolome profiling technologies are rapidly advancing with increasing coverage of thousands of proteins and hundreds of metabolites, although they remain far from comprehensive. Due to technical limitations, affinity-based proteome profiling datasets can only be used to compare one protein across samples, not between proteins among different samples, which limits the value of the data. Currently, the metabolomic profiling is conducted solely using MS, which provides very sensitive, accurate, and reproducible data, however, is not readily available to most studies and restricts its utility. Proteomic and metabolomic profiling require both technological and analytical advancements to maximize their clinical applications.

Multi-omics profiling offers significant potential to revolutionize clinical practice. Considering that processing and interpreting complex omics data presents a major challenge, robust computational strategies, such as machine learning algorithms, are evolving rapidly [257,258]. Deep learning, a prominent category of ML algorithms that has gained considerable attention in recent years, mirrors the neurological framework of the brain. By learning from multi-layered neural network architectures, it is capable of identifying complex patterns and making predictions. These models extract high-level features from input data by processing them across various modalities. Consequently, the integration of clinical data, histopathological images, and different types of omics data - including genomic, transcriptomic, and methylome profiles - during training can significantly enhance the predictive power of these models.

A wide range of deep learning algorithms have been developed to aim at improving patient diagnosis, prognosis, and treatment management. For example, Islam et al. developed a multimodal framework that improved the prediction of breast cancer subtypes by employing two convolutional neural networks (CNNs) models trained with CNV and transcriptome data respectively [259]. Similarly, Foersch et al. developed a deep learning tool for characterizing anti-tumor immunity that outperformed other clinical, molecular, and immune cell-based parameters in predicting prognosis and therapy response in colorectal cancer patients [260]. Despite facing challenges such as difficulties in training

Table 2
Examples of clinical application of omics signatures in cancer.

Type of signature	Type of cancer	Clinical application	Reference
Mutational signature	Colorectal cancer	Biomarkers of biallelic *MUTYH* germline variants	[47,48]
	Endometrial cancer, colorectal cancer	Biomarker of polymerase proofreading and mismatch repair defective tumors (*POLE*, *POLD1* deficiency)	[268]
	Clear cell renal cell carcinoma	Biomarker of *ERCC2* deficiency	[269]
	Breast and ovarian cancers	Responses to PARP inhibitors	[58,270]
	Ovarian, pancreatic, gastrointestinal (GI) and thoracic cancers	Responses to platinum-based therapy	[20,61, 271,272]
	Pan-cancer	Biomarker of BRCA-deficiency, responses to ATR inhibitor	[273]
	Pan-cancer	Biomarker of mismatch repair deficiency, responses to immune-checkpoint-inhibitor	[64]
Methylation signature	Central nervous system cancer and sarcoma	Subtype classification	[81,93]
	Colorectal and lung cancer	Diagnosis and early detection	[104,105]
	Pan-cancer	Diagnosis, early detection and determination of tissue of origin	[108,274]
	Bladder cancer	Monitoring for relapse	[113]
	Breast cancer, lung adenocarcinoma and neuroblastoma	Prognosis	[116,275, 276]
	Gastric and prostate cancer	Monitoring for metastasis	[122–124]
	Non–small-cell lung carcinoma and melanoma	Responses to immune-checkpoint-inhibitor	[135,277, 278]
	Non-small cell lung cancer	Responses to cisplatin	[132]
Transcriptome signature	Carcinoma with unknown primary	Determination of tissue of origin	[154,155]
	Colorectal, bladder and breast cancer	Subtype classification	[156,158, 279]
	Oral squamous cell cancer	Diagnosis and early detection	[163,168]
	Lung adenocarcinoma, colorectal and breast cancer	Prognosis and prediction of metastasis	[169,172, 280]
	Leukemia	Treatment decision	[175,176]
	Melanoma and lung adenocarcinoma	Responses to immune-checkpoint-inhibitor	[281,282]
Proteomic and metabolic signatures	Ovarian cancer	Early detection	[204]
	Liver cancer, pancreatic cancer	Subtype classification	[189,191]
	Melanoma	Responses to immune-checkpoint-inhibitor	[237]
	Lung cancer	Predict resistance to chemotherapy	[239]

and interpreting the trained model, researchers are actively working on developing more efficient training methods and explainable algorithms. These efforts aim to overcome these challenges and enhance the future applicability of deep learning in clinical cancer applications.

7. Conclusion and directions for future research

In this minireview, we have summarized the development of omics signatures as well as their current and potential clinical applications in cancer (Table 2). The pivotal role of these signatures in disease progression and their value as diagnostic and prognostic markers have been emphasized. The availability of a large amount of genome-wide sequencing data has provided opportunities for understanding the link between phenotypes and their molecular underpinnings, thereby aiding the clinical decision-making process. These findings should be integrated with clinical data to enhance our understanding of the etiology and disease association of these signatures. The challenge lies in analyzing and interpreting the data, identifying useful signatures within the vast amount of data, and converting these into actionable information for clinical application. Moreover, clinical validation of the developed omics signature classifiers remains essential for the translation from bench-side to bedside. This can be particularly highlighted by the success of the DKFZ/Heidelberg CNS tumor classifier which now is incorporated into the WHO 2021 CNS tumor classification guideline [81,92]. Similar efforts have been made such as the setup of the Circulating Cell-free Genome Atlas study for proving the clinical value of methylation signature in cancer diagnosis [108,110]. Future efforts should focus on the development of computational tools and algorithms to accurately summarize and infer clinically relevant signatures. Moving forward, the adoption of multi-omics data in the precision medicine framework is the trend. Integration of the large amounts of multi-omic and clinical data would provide new insights and potentially revolutionize the landscape of clinical management, including early detection and classification of diseases, as well as development of personalized therapies. More efforts will be dedicated to achieving more comprehensive analysis, interpretation, and visualization of multi-omics data.

CRediT authorship contribution statement

Wei Ma: Conceptualization, Writing - original draft, Writing – review & editing, Visualization. **Wenshu Tang:** Conceptualization, Writing – original draft, Writing – review & editing, Visualization. **Jamie S.L. Kwok:** Conceptualization, Writing – original draft. **Amy H. Y. Tong:** Writing – review & editing. **Cario W.S. Lo:** Writing – review & editing. **Annie T.W. Chu:** Writing – review & editing, Supervision. Brian **H.Y. Chung:** Conceptualization, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Dr Su-vui Lo, Chief Executive Officer, of the Hong Kong Genome Institute for the instrumental leadership and guidance.

References

[1] Koh G, Degasperi A, Zou X, Momen S, Nik-Zainal S. Mutational signatures: emerging concepts, caveats and clinical applications. Nat Rev Cancer 2021;21: 619–37. https://doi.org/10.1038/s41568-021-00377-7.
[2] Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature 2013;500: 415–21. https://doi.org/10.1038/nature12477.
[3] Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nat Med 2017;23:517–25. https://doi.org/10.1038/nm.4292.
[4] Smid M, Rodríguez-González FG, Sieuwerts AM, Salgado R, Smissen WJCP, et al. Breast cancer genome and transcriptome integration implicates specific

mutational signatures with immune cell infiltration. Nat Commun 2016;7:12910. https://doi.org/10.1038/ncomms12910.
[5] Zhang Z, Chen L, Chen H, Zhao J, Li K, Sun J, et al. Pan-cancer landscape of T-cell exhaustion heterogeneity within the tumor microenvironment revealed a progressive roadmap of hierarchical dysfunction associated with prognosis and therapeutic efficacy. EBioMedicine 2022;83:104207. https://doi.org/10.1016/j.ebiom.2022.104207.
[6] Nirmal AJ, Regan T, Shih BB, Hume DA, Sims AH, Freeman TC. Immune cell gene signatures for profiling the microenvironment of solid tumors. canimm.0342.2018 Cancer Immunol Res 2018;6. https://doi.org/10.1158/2326-6066.cir-18-0342.
[7] Omichessan H, Severi G, Perduca V. Computational tools to detect signatures of mutational processes in DNA from tumours: A review and empirical comparison of performance. PLoS ONE 2019;14:e0221235. https://doi.org/10.1371/journal.pone.0221235.
[8] Rauschert S, Raubenheimer K, Melton PE, Huang RC. Machine learning and clinical epigenetics: a review of challenges for diagnosis and classification. Clin Epigenetics 2020;12:51. https://doi.org/10.1186/s13148-020-00842-4.
[9] Azad A, Pavlopoulos GA, Ouzounis CA, Kyrpides NC, Buluç A. HipMCL: a high-performance parallel implementation of the Markov clustering algorithm for large-scale networks. gkx1313- Nucleic Acids Res 2018;46. https://doi.org/10.1093/nar/gkx1313.
[10] Nik-Zainal S, Alexandrov LB, Wedge DC, Loo PV, Greenman CD, Raine K, et al. Mutational Processes Molding the Genomes of 21 Breast Cancers. Cell 2012;149:979–93. https://doi.org/10.1016/j.cell.2012.04.024.
[11] Chang K, Creighton CJ, Davis C, Donehower L, Drummond J, Wheeler D, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 2013;45:1113–20. https://doi.org/10.1038/ng.2764.
[12] Consortium TIP-CA of WG, Aaltonen LA, Abascal F, Abeshouse A, Aburatani H, Adams DJ, et al. Pan-cancer analysis of whole genomes. Nature 2020;578:82–93. https://doi.org/10.1038/s41586-020-1969-6.
[13] Morganella S, Alexandrov LB, Glodzik D, Zou X, Davies H, Staaf J, et al. The topography of mutational processes in breast cancer genomes. Nat Commun 2016;7:11383. https://doi.org/10.1038/ncomms11383.
[14] Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 2016;534:47–54. https://doi.org/10.1038/nature17676.
[15] Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Ng AWT, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature 2020;578:94–101. https://doi.org/10.1038/s41586-020-1943-3.
[16] Degasperi A, Zou X, Amarante TD, Martinez-Martinez A, Koh GCC, Dias JML, et al. Substitution mutational signatures in whole-genome–sequenced cancers in the UK population. Science 2022;376. https://doi.org/10.1126/science.abl9283.
[17] Singh VK, Rastogi A, Hu X, Wang Y, De S. Mutational signature SBS8 predominantly arises due to late replication errors in cancer. Commun Biol 2020;3:421. https://doi.org/10.1038/s42003-020-01119-5.
[18] Polak P, Kim J, Braunstein LZ, Karlic R, Haradhavala NJ, Tiao G, et al. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. Nat Genet 2017;49:1476–86. https://doi.org/10.1038/ng.3934.
[19] Fang H, Zhu X, Yang H, Oh J, Barbour JA, Wong JWH. Deficiency of replication-independent DNA mismatch repair drives a 5-methylcytosine deamination mutational signature in cancer. Sci Adv 2021;7:eabg4398. https://doi.org/10.1126/sciadv.abg4398.
[20] Jager M, Blokzijl F, Kuijk E, Bertl J, Vougioukalaki M, Janssen R, et al. Deficiency of nucleotide excision repair is associated with mutational signature observed in cancer. Genome Res 2019;29:1067–77. https://doi.org/10.1101/gr.246223.118.
[21] Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. Cell Rep 2013;3:246–59. https://doi.org/10.1016/j.celrep.2012.12.008.
[22] Ardin M, Cahais V, Castells X, Bouaoun L, Byrnes G, Herceg Z, et al. MutSpec: a Galaxy toolbox for streamlined analyses of somatic mutation spectra in human and mouse cancer genomes. BMC Bioinform 2016;17:170. https://doi.org/10.1186/s12859-016-1011-z.
[23] Blokzijl F, Janssen R, Boxtel R van, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med 2018;10:33. https://doi.org/10.1186/s13073-018-0539-0.
[24] Degasperi A, Amarante TD, Czarnecki J, Shooter S, Zou X, Glodzik D, et al. A practical framework and online tool for mutational signature analyses show intertissue variation and driver dependencies. Nat Cancer 2020;1:249–63. https://doi.org/10.1038/s43018-020-0027-5.
[25] Mayakonda A, Lin D-C, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res 2018;28:1747–56. https://doi.org/10.1101/gr.239244.118.
[26] Wang S, Li H, Song M, Tao Z, Wu T, He Z, et al. Copy number signature analysis tool and its application in prostate cancer reveals distinct mutational processes and clinical outcomes. PLoS Genet 2021;17:e1009557. https://doi.org/10.1371/journal.pgen.1009557.
[27] Gehring JS, Fischer B, Lawrence M, Huber W. SomaticSignatures: inferring mutational signatures from single-nucleotide variants. Bioinformatics 2015;31:3673–5. https://doi.org/10.1093/bioinformatics/btv408.
[28] Islam SMA, Díaz-Gay M, Wu Y, Barnes M, Vangara R, Bergstrom EN, et al. Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. Cell Genom 2022;2:100179. https://doi.org/10.1016/j.xgen.2022.100179.
[29] Díaz-Gay M, Vangara R, Barnes M, Wang X, Islam SMA, Vermes I, et al. Assigning mutational signatures to individual samples and individual somatic mutations with SigProfilerAssignment. 2023.07.10.548264 BioRxiv 2023. https://doi.org/10.1101/2023.07.10.548264.
[30] Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol 2016;17:31. https://doi.org/10.1186/s13059-016-0893-4.
[31] Afsari B, Kuo A, Zhang Y, Li L, Lahouel K, Danilova L, et al. Supervised mutational signatures for obesity and other tissue-specific etiological factors in cancer. ELife 2021;10:e61082. https://doi.org/10.7554/elife.61082.
[32] Maura F, Degasperi A, Nadeu F, Leongamornlert D, Davies H, Moore L, et al. A practical guide for mutational signature analysis in hematological malignancies. Nat Commun 2019;10:2969. https://doi.org/10.1038/s41467-019-11037-8.
[33] Li Y, Roberts ND, Wala JA, Shapira O, Schumacher SE, Kumar K, et al. Patterns of somatic structural variation in human cancer genomes. Nature 2020;578:112–21. https://doi.org/10.1038/s41586-019-1913-9.
[34] Maclachlan KH, Rustad EH, Derkach A, Zheng-Lin B, Yellapantula V, Diamond B, et al. Copy number signatures predict chromothripsis and clinical outcomes in newly diagnosed multiple myeloma. Nat Commun 2021;12:5172. https://doi.org/10.1038/s41467-021-25469-8.
[35] Hadi K, Yao X, Behr JM, Deshpande A, Xanthopoulakis C, Tian H, et al. Distinct Classes of Complex Structural Variation Uncovered across Thousands of Cancer Genome Graphs. Cell 2020;183:197–210.e32. https://doi.org/10.1016/j.cell.2020.08.006.
[36] Steele CD, Abbasi A, Islam SMA, Bowes AL, Khandekar A, Haase K, et al. Signatures of copy number alterations in human cancer. Nature 2022;606:984–91. https://doi.org/10.1038/s41586-022-04738-6.
[37] Macintyre G, Goranova TE, Silva DD, Ennis D, Piskorz AM, Eldridge M, et al. Copy number signatures and mutational processes in ovarian carcinoma. Nat Genet 2018;50:1262–70. https://doi.org/10.1038/s41588-018-0179-8.
[38] Drews RM, Hernando B, Tarabichi M, Haase K, Lesluyes T, Smith PS, et al. A pan-cancer compendium of chromosomal instability. Nature 2022;606:976–83. https://doi.org/10.1038/s41586-022-04789-9.
[39] Bergstrom EN, Huang MN, Mahto U, Barnes M, Stratton MR, Rozen SG, et al. SigProfilerMatrixGenerator: a tool for visualizing and exploring patterns of small mutational events. BMC Genom 2019;20:685. https://doi.org/10.1186/s12864-019-6041-2.
[40] Khandekar A, Vangara R, Barnes M, Díaz-Gay M, Abbasi A, Bergstrom EN, et al. Visualizing and exploring patterns of large mutational events with SigProfilerMatrixGenerator. BMC Genom 2023;24:469. https://doi.org/10.1186/s12864-023-09584-y.
[41] Thutkawkorapin J, Eisfeldt J, Tham E, Nilsson D. pyCancerSig: subclassifying human cancer with comprehensive single nucleotide, structural and microsatellite mutational signature deconstruction from whole genome sequencing. BMC Bioinform 2020;21:128. https://doi.org/10.1186/s12859-020-3451-8.
[42] Sugita I, Matsuyama S, Dobashi H, Komura D, Ishikawa S. Viola: a structural variant signature extractor with user-defined classifications. btab662 Bioinformatics 2021;38. https://doi.org/10.1093/bioinformatics/btab662.
[43] Bao L, Zhong X, Yang Y, Yang L. Starfish infers signatures of complex genomic rearrangements across human cancers. Nat Cancer 2022;3:1247–59. https://doi.org/10.1038/s43018-022-00404-y.
[44] Rahman N. Realizing the promise of cancer predisposition genes. Nature 2014;505:302–8. https://doi.org/10.1038/nature12981.
[45] Esteller M. Epigenetic lesions causing genetic lesions in human cancer promoter hypermethylation of DNA repair genes. Eur J Cancer 2000;36:2294–300. https://doi.org/10.1016/s0959-8049(00)00303-8.
[46] Knijnenburg TA, Wang L, Zimmermann MT, Chambwe N, Gao GF, Cherniack AD, et al. Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. Cell Rep 2018;23:239–254.e6. https://doi.org/10.1016/j.celrep.2018.03.076.
[47] Georgeson P, Pope BJ, Rosty C, Clendenning M, Mahmood K, Joo JE, et al. Evaluating the utility of tumour mutational signatures for identifying hereditary colorectal cancer and polyposis syndrome carriers. Gut 2021;70:2138–49. https://doi.org/10.1136/gutjnl-2019-320462.
[48] Georgeson P, Harrison TA, Pope BJ, Zaidi SH, Qu C, Steinfelder RS, et al. Identifying colorectal cancer caused by biallelic MUTYH pathogenic variants using tumor mutational signatures. Nat Commun 2022;13:3254. https://doi.org/10.1038/s41467-022-30916-1.
[49] Grolleman, Voer JE, de RM, Elsayed FA, Nielsen M, Weren RDA, Palles C, et al. Mutational Signature Analysis Reveals NTHL1 Deficiency to Cause a Multi-tumor Phenotype. Cancer Cell 2019;35:256–266.e5. https://doi.org/10.1016/j.ccell.2018.12.011.
[50] Sanjaya P, Maljanen K, Katainen R, Waszak SM, Ambrose JC, Arumugam P, et al. Mutation-Attention (MuAt): deep representation learning of somatic mutations for tumour typing and subtyping. Genome Med 2023;15:47. https://doi.org/10.1186/s13073-023-01204-4.
[51] Moon I, LoPiccolo J, Baca SC, Sholl LM, Kehl KL, Hassett MJ, et al. Machine learning for genetics-based classification and treatment response prediction in cancer of unknown primary. Nat Med 2023;29:2057–67. https://doi.org/10.1038/s41591-023-02482-6.
[52] Saito-Adachi M, Hama N, Totoki Y, Nakamura H, Arai Y, Hosoda F, et al. Oncogenic structural aberration landscape in gastric cancer genomes. Nat Commun 2023;14:3688. https://doi.org/10.1038/s41467-023-39263-1.

[53] Yang Y, Yang L. Somatic structural variation signatures in pediatric brain tumors. 2023.05.18.23290139 MedRxiv 2023. https://doi.org/10.1101/2023.05.18.23290139.
[54] Hopkins JL, Lan L, Zou L. DNA repair defects in cancer and therapeutic opportunities. Genes Dev 2022;36:278–93. https://doi.org/10.1101/gad.349431.122.
[55] Javle M, Curtin NJ. The role of PARP in DNA repair and its therapeutic exploitation. Br J Cancer 2011;105:1114–22. https://doi.org/10.1038/bjc.2011.382.
[56] Hsiehchen D, Hsieh A, Samstein RM, Lu T, Beg MS, Gerber DE, et al. DNA Repair Gene Mutations as Predictors of Immune Checkpoint Inhibitor Response beyond Tumor Mutation Burden. Cell Rep Med 2020;1:100034. https://doi.org/10.1016/j.xcrm.2020.100034.
[57] Pilger D, Seymour LW, Jackson SP. Interfaces between cellular responses to DNA damage and cancer immunotherapy. Genes Dev 2021;35:602–18. https://doi.org/10.1101/gad.348314.121.
[58] Batalini F, Gulhan DC, Mao V, Tran A, Polak M, Xiong N, et al. Mutational Signature 3 Detected from Clinical Panel Sequencing is Associated with Responses to Olaparib in Breast and Ovarian Cancers. OF1–10 Clin Cancer Res 2022;28. https://doi.org/10.1158/1078-0432.ccr-22-0749.
[59] Alexandrov LB, Nik-Zainal S, Siu HC, Leung SY, Stratton MR. A mutational signature in gastric cancer suggests therapeutic strategies. Nat Commun 2015;6:8683. https://doi.org/10.1038/ncomms9683.
[60] Gulhan DC, Lee JJ-K, Melloni GEM, Cortés-Ciriano I, Park PJ. Detecting the mutational signature of homologous recombination deficiency in clinical samples. Nat Genet 2019;51:912–9. https://doi.org/10.1038/s41588-019-0390-2.
[61] Zhao EY, Shen Y, Pleasance E, Kasaian K, Leelakumari S, Jones M, et al. Homologous Recombination Deficiency and Platinum-Based Therapy Outcomes in Advanced Breast Cancer. Clin Cancer Res 2017;23:7521–30. https://doi.org/10.1158/1078-0432.ccr-17-1941.
[62] Chopra N, Tovey H, Pearson A, Cutts R, Toms C, Proszek P, et al. Homologous recombination DNA repair deficiency and PARP inhibition activity in primary triple negative breast cancer. Nat Commun 2020;11:2662. https://doi.org/10.1038/s41467-020-16142-7.
[63] Cherri S, Oneda E, Noventa S, Melocchi L, Zaniboni A. Microsatellite instability and chemosensitivity in solid tumours. 17588359221099348 Ther Adv Méd Oncol 2022;14. https://doi.org/10.1177/17588359221099347.
[64] Bever KM, Le DT. DNA repair defects and implications for immunotherapy. J Clin Invest 2018;128:4236–42. https://doi.org/10.1172/jci122010.
[65] Zou X, Koh GCC, Nanda AS, Degasperi A, Urgo K, Roumeliotis TI, et al. A systematic CRISPR screen defines mutational mechanisms underpinning signatures caused by replication errors and endogenous DNA damage. Nat Cancer 2021;2:643–57. https://doi.org/10.1038/s43018-021-00200-0.
[66] Levatić J, Salvadores M, Fuster-Tormo F, Supek F. Mutational signatures are markers of drug sensitivity of cancer cells. Nat Commun 2022;13:2926. https://doi.org/10.1038/s41467-022-30582-3.
[67] Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol 2019;20:590–607. https://doi.org/10.1038/s41580-019-0159-6.
[68] Watt F, Molloy PL. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev 1988;2:1136–43. https://doi.org/10.1101/gad.2.9.1136.
[69] Ben-Hattar J, Jiricny J. Methylation of single CpG dinucleotides within a promoter element of the Herpes simplex virus tk gene reduces its transcription in vivo. Gene 1988;65:219–27. https://doi.org/10.1016/0378-1119(88)90458-1.
[70] Nishiyama A, Nakanishi M. Navigating the DNA methylation landscape of cancer. Trends Genet 2021;37:1012–27. https://doi.org/10.1016/j.tig.2021.05.002.
[71] Locke WJ, Guanzon D, Ma C, Liew YJ, Duesing KR, Fung KYC, et al. DNA Methylation Cancer Biomarkers: Translation to the Clinic. Front Genet 2019;10:1150. https://doi.org/10.3389/fgene.2019.01150.
[72] Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov 2022;12:31–46. https://doi.org/10.1158/2159-8290.cd-21-1059.
[73] Zhao SG, Chen WS, Li H, Foye A, Zhang M, Sjöström M, et al. The DNA methylation landscape of advanced prostate cancer. Nat Genet 2020;52:778–89. https://doi.org/10.1038/s41588-020-0648-8.
[74] Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, et al. Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. Cancer Cell 2010;18:553–67. https://doi.org/10.1016/j.ccr.2010.11.015.
[75] Costoya JA, Arce VM. Cancer cells escape the immune system by increasing stemness through epigenetic reprogramming. Cell Mol Immunol 2023;20:6–7. https://doi.org/10.1038/s41423-022-00953-3.
[76] Suzuki M, Liao W, Wos F, Johnston AD, DeGrazia J, Ishii J, et al. Whole-genome bisulfite sequencing with improved accuracy and cost. Genome Res 2018;28:1364–71. https://doi.org/10.1101/gr.232587.117.
[77] Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol 2016;17:208. https://doi.org/10.1186/s13059-016-1066-1.
[78] Sadikovic B, Levy MA, Kerkhof J, Aref-Eshghi E, Schenkel L, Stuart A, et al. Clinical epigenomics: genome-wide DNA methylation analysis for the diagnosis of Mendelian disorders. Genet Med 2021;23:1065–74. https://doi.org/10.1038/s41436-020-01096-4.
[79] Galbraith K, Vasudevaraja V, Serrano J, Shen G, Tran I, Abdallat N, et al. Clinical utility of whole-genome DNA methylation profiling as a primary molecular diagnostic assay for central nervous system tumors—A prospective study and guidelines for clinical testing. Neuro-Oncol Adv 2023;5:vdad076. https://doi.org/10.1093/noajnl/vdad076.
[80] Haghshenas S, Bhai P, Aref-Eshghi E, Sadikovic B. Diagnostic Utility of Genome-Wide DNA Methylation Analysis in Mendelian Neurodevelopmental Disorders. Int J Mol Sci 2020;21:9303. https://doi.org/10.3390/ijms21239303.
[81] Capper D, Jones DTW, Sill M, Hovestadt V, Schrimpf D, Sturm D, et al. DNA methylation-based classification of central nervous system tumours. Nature 2018;555:469–74. https://doi.org/10.1038/nature26000.
[82] Aref-Eshghi E, Rodenhiser DI, Schenkel LC, Lin H, Skinner C, Ainsworth P, et al. Genomic DNA methylation signatures enable concurrent diagnosis and clinical genetic variant classification in neurodevelopmental syndromes. Am J Hum Genet 2018;102:156–74. https://doi.org/10.1016/j.ajhg.2017.12.008.
[83] Pienkowska M, Choufani S, Turinsky AL, Guha T, Merino DM, Novokmet A, et al. DNA methylation signature is prognostic of choroid plexus tumor aggressiveness. Clin Epigenetics 2019;11:117. https://doi.org/10.1186/s13148-019-0708-z.
[84] Aref-Eshghi E, Kerkhof J, Pedro VP, France GD, Barat-Houari M, Ruiz-Pallares N, et al. Evaluation of DNA methylation episignatures for diagnosis and phenotype correlations in 42 mendelian neurodevelopmental disorders. Am J Hum Genet 2020;106:356–70. https://doi.org/10.1016/j.ajhg.2020.01.019.
[85] Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinform 2012;13:86. https://doi.org/10.1186/1471-2105-13-86.
[86] Levy MA, McConkey H, Kerkhof J, Barat-Houari M, Bargiacchi S, Biamino E, et al. Novel diagnostic DNA methylation episignatures expand and refine the epigenetic landscapes of Mendelian disorders. Hum Genet Genom Adv 2022;3:100075. https://doi.org/10.1016/j.xhgg.2021.100075.
[87] Li W, Cerise JE, Yang Y, Han H. Application of t-SNE to human genetic data. J Bioinform Comput Biol 2017;15:1750017. https://doi.org/10.1142/s0219720017500172.
[88] Dorrity MW, Saunders LM, Queitsch C, Fields S, Trapnell C. Dimensionality reduction by UMAP to visualize physical and genetic interactions. Nat Commun 2020;11:1537. https://doi.org/10.1038/s41467-020-15351-4.
[89] Maros ME, Capper D, Jones DTW, Hovestadt V, Deimling A von, Pfister SM, et al. Machine learning workflows to estimate class probabilities for precision cancer diagnostics on DNA methylation microarray data. Nat Protoc 2020;15:479–512. https://doi.org/10.1038/s41596-019-0251-6.
[90] Northcott PA, Korshunov A, Witt H, Hielscher T, Eberhart CG, Mack S, et al. Medulloblastoma Comprises Four Distinct Molecular Variants. J Clin Oncol 2010;29:1408–14. https://doi.org/10.1200/jco.2009.27.4324.
[91] Hovestadt V, Remke M, Kool M, Pietsch T, Northcott PA, Fischer R, et al. Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density DNA methylation arrays. Acta Neuropathol 2013;125(6):913. https://doi.org/10.1007/s00401-013-1126-5.
[92] Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, et al. The 2021 WHO classification of tumors of the central nervous system: a summary. Neuro-Oncol 2021;23:1231–51. https://doi.org/10.1093/neuonc/noab106.
[93] Koelsche C, Schrimpf D, Stichel D, Sill M, Sahm F, Reuss DE, et al. Sarcoma classification by DNA methylation profiling. Nat Commun 2021;12:498. https://doi.org/10.1038/s41467-020-20603-4.
[94] Gao F, Wang W, Tan M, Zhu L, Zhang Y, Fessler E, et al. DeepCC: a novel deep learning-based framework for cancer molecular subtype classification. Oncogenesis 2019;8:44. https://doi.org/10.1038/s41389-019-0157-8.
[95] Zhang S, Wang Y, Gu Y, Zhu J, Ci C, Guo Z, et al. Specific breast cancer prognosis-subtype distinctions based on DNA methylation patterns. Mol Oncol 2018;12:1047–60. https://doi.org/10.1002/1878-0261.12309.
[96] Filipski K, Scherer M, Zeiner KN, Bucher A, Kleemann J, Jurmeister P, et al. DNA methylation-based prediction of response to immune checkpoint inhibition in metastatic melanoma. J Immunother Cancer 2021;9:e002226. https://doi.org/10.1136/jitc-2020-002226.
[97] Li M, Zhu C, Xue Y, Miao C, He R, Li W, et al. A DNA methylation signature for the prediction of tumour recurrence in stage II colorectal cancer. Br J Cancer 2023;128(9):1681. https://doi.org/10.1038/s41416-023-02155-8.
[98] Peng Y, Wu Q, Wang L, Wang H, Yin F. A DNA methylation signature to improve survival prediction of gastric cancer. Clin Epigenetics 2020;12:15. https://doi.org/10.1186/s13148-020-0807-x.
[99] Petko Z, Ghiassi M, Shuber A, Gorham J, Smalley W, Washington MK, et al. Aberrantly methylated CDKN2A, MGMT, and MLH1 in colon polyps and in fecal DNA from patients with colorectal polyps. Clin Cancer Res: J Am Assoc Cancer Res 2005;11:1203–9.
[100] Kim MS, Lee J, Sidransky D. DNA methylation markers in colorectal cancer. Cancer Metastas– Rev 2010;29:181–206. https://doi.org/10.1007/s10555-010-9207-6.
[101] IT F, RD F, IS H, LT R, L Philip, LG P, et al. Multitarget stool DNA testing for colorectal-cancer screening. N Engl J Med 2014;370:1287–97. https://doi.org/10.1056/nejmoa1311194.
[102] Hernandez-Meza G, Felden J, Gonzalez-Kozlova EE, Garcia-Lezana T, Peix J, Portela A, et al. DNA methylation profiling of human hepatocarcinogenesis. Hepatology 2021;74:183–99. https://doi.org/10.1002/hep.31659.
[103] Taryma-Leśniak O, Sokolowska KE, Wojdacz TK. Current status of development of methylation biomarkers for in vitro diagnostic IVD applications. Clin Epigenetics 2020;12:100. https://doi.org/10.1186/s13148-020-00886-6.
[104] Shirley M. Epi proColon® for colorectal cancer screening: a profile of its use in the USA. Mol Diagn Ther 2020;24:497–503. https://doi.org/10.1007/s40291-020-00473-8.

[105] Dietrich D, Kneip C, Raji O, Liloglou T, Seegebarth A, Schlegel T, et al. Performance evaluation of the DNA methylation biomarker SHOX2 for the aid in diagnosis of lung cancer based on the analysis of bronchial aspirates. Int J Oncol 2011;40:825–32. https://doi.org/10.3892/ijo.2011.1264.
[106] 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;359:926–30. https://doi.org/10.1126/science.aar3247.
[107] Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV, Consortium C, et al. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. Ann Oncol 2020;31:745–59. https://doi.org/10.1016/j.annonc.2020.02.011.
[108] Klein EA, Richards D, Cohn A, Tummala M, Lapham R, Cosgrove D, et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. Ann Oncol 2021;32:1167–77. https://doi.org/10.1016/j.annonc.2021.05.806.
[109] Stackpole ML, Zeng W, Li S, Liu C-C, Zhou Y, He S, et al. Cost-effective methylome sequencing of cell-free DNA for accurately detecting and locating cancer. Nat Commun 2022;13:5566. https://doi.org/10.1038/s41467-022-32995-6.
[110] Jamshidi A, Liu MC, Klein EA, Venn O, Hubbell E, Beausang JF, et al. Evaluation of cell-free DNA approaches for multi-cancer early detection. e12 Cancer Cell 2022;40:1537–49. https://doi.org/10.1016/j.ccell.2022.10.022.
[111] Nicholson BD, Oke J, Virdee PS, Harris DA, O'Doherty C, Park JE, et al. Multi-cancer early detection test in symptomatic patients referred for cancer investigation in England and Wales (SYMPLIFY): a large-scale, observational cohort study. Lancet Oncol 2023;24:733–43. https://doi.org/10.1016/s1470-2045(23)00277-2.
[112] Witjes, Morote JA, Cornel J, Gakis EB, Valenberg G, van FJP, Lozano F, et al. Performance of the Bladder EpiCheck™ methylation test for patients under surveillance for non–muscle-invasive bladder cancer: results of a multicenter, prospective, blinded clinical trial. Eur Urol Oncol 2018;1:307–13. https://doi.org/10.1016/j.euo.2018.06.011.
[113] D'Andrea D, Soria F, Zehetmayer S, Gust KM, Korn S, Witjes JA, et al. Diagnostic accuracy, clinical utility and influence on decision-making of a methylation urine biomarker test in the surveillance of non-muscle-invasive bladder cancer. BJU Int 2019;123:959–67. https://doi.org/10.1111/bju.14673.
[114] Young GP, Pedersen SK, Mansfield S, Murray DH, Baker RT, Rabbitt P, et al. A cross-sectional study comparing a blood test for methylated BCAT1 and IKZF1 tumor-derived DNA with CEA for detection of recurrent colorectal cancer. Cancer Med 2016;5:2763–72. https://doi.org/10.1002/cam4.868.
[115] Symonds EL, Pedersen SK, Yeo B, Naji HA, Byrne SE, Roy A, et al. Assessment of tumor burden and response to therapy in patients with colorectal cancer using a quantitative ctDNA test for methylated BCAT1/IKZF1. Mol Oncol 2022;16:2031–41. https://doi.org/10.1002/1878-0261.13178.
[116] Jeschke J, Bizet M, Desmedt C, Calonne E, Dedeurwaerder S, Garaud S, et al. DNA methylation–based immune response signature improves patient diagnosis in multiple cancers. J Clin Invest 2017;127:3090–102. https://doi.org/10.1172/jci91095.
[117] Batra RN, Lifshitz A, Vidakovic AT, Chin S-F, Sati-Batra A, Sammut S-J, et al. DNA methylation landscapes of 1538 breast cancers reveal a replication-linked clock, epigenomic instability and cis-regulation. Nat Commun 2021;12:5406. https://doi.org/10.1038/s41467-021-25661-w.
[118] Villanueva A, Portela A, Sayols S, Battiston C, Hoshida Y, Méndez-González J, et al. DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. Hepatology 2015;61:1945–56. https://doi.org/10.1002/hep.27732.
[119] Henriksen SD, Madsen PH, Larsen AC, Johansen MB, Pedersen IS, Krarup H, et al. Promoter hypermethylation in plasma-derived cell-free DNA as a prognostic marker for pancreatic adenocarcinoma staging. Int J Cancer 2017;141:2489–97. https://doi.org/10.1002/ijc.31024.
[120] Jin S, Zhu D, Shao F, Chen S, Guo Y, Li K, et al. Efficient detection and post-surgical monitoring of colon cancer with a multi-marker DNA methylation liquid biopsy. Proc Natl Acad Sci 2021;118:e2017421118. https://doi.org/10.1073/pnas.2017421118.
[121] Herrgott GA, Snyder JM, She R, Malta TM, Sabedot TS, Lee IY, et al. Detection of diagnostic and prognostic methylation-based signatures in liquid biopsy specimens from patients with meningiomas. Nat Commun 2022;14:5669. https://doi.org/10.1038/s41467-023-41434-z.
[122] Wu J, Xiao Y, Xia C, Yang F, Li H, Shao Z, et al. Identification of Biomarkers for Predicting Lymph Node Metastasis of Stomach Cancer Using Clinical DNA Methylation Data. Dis Markers 2017;2017:5745724. https://doi.org/10.1155/2017/5745724.
[123] Chen S, Yu Y, Li T, Ruan W, Wang J, Peng Q, et al. A novel DNA methylation signature associated with lymph node metastasis status in early gastric cancer. Clin Epigenetics 2022;14:18. https://doi.org/10.1186/s13148-021-01219-x.
[124] Wu A, Cremaschi P, Wetterskog D, Conteduca V, Franceschini GM, Kleftogiannis D, et al. Genome-wide plasma DNA methylation features of metastatic prostate cancer. J Clin Invest 2020;130:1991–2000. https://doi.org/10.1172/jci130887.
[125] Rodger EJ, Gimenez G, Ajithkumar P, Stockwell PA, Almomani S, Bowden SA, et al. An epigenetic signature of advanced colorectal cancer metastasis. IScience 2023;26:106986. https://doi.org/10.1016/j.isci.2023.106986.
[126] Li W, Guo L, Tang W, Ma Y, Wang X, Shao Y, et al. Identification of DNA methylation biomarkers for risk of liver metastasis in early-stage colorectal cancer. Clin Epigenetics 2021;13:126. https://doi.org/10.1186/s13148-021-01108-3.
[127] Xu Y, Huang Z, Yu X, Chen K, Fan Y. Integrated genomic and DNA methylation analysis of patients with advanced non-small cell lung cancer with brain metastases. Mol Brain 2021;14:176. https://doi.org/10.1186/s13041-021-00886-4.
[128] Hainsworth JD, Rubin MS, Spigel DR, Boccia RV, Raby S, Quinn R, et al. Molecular Gene Expression Profiling to Predict the Tissue of Origin and Direct Site-Specific Therapy in Patients With Carcinoma of Unknown Primary Site: A Prospective Trial of the Sarah Cannon Research Institute. J Clin Oncol 2012;31:217–23. https://doi.org/10.1200/jco.2012.43.3755.
[129] Moran S, Martínez-Cardús A, Sayols S, Musulén E, Balañá C, Estival-Gonzalez A, et al. Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. Lancet Oncol 2016;17:1386–95. https://doi.org/10.1016/s1470-2045(16)30297-2.
[130] HM E, Annie-Claire D, Thierry G, Marie-France H, Nicolas de T, Michael W, et al. MGMT Gene Silencing and Benefit from Temozolomide in Glioblastoma. N Engl J Med 2005;352:997–1003. https://doi.org/10.1056/nejmoa043331.
[131] Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, et al. A landscape of pharmacogenomic interactions in cancer. Cell 2016;166:740–54. https://doi.org/10.1016/j.cell.2016.06.017.
[132] Zhang Y-W, Zheng Y, Wang J-Z, LU X-X, Wang Z, Chen L-B, et al. Integrated analysis of DNA methylation and mRNA expression profiling reveals candidate genes associated with cisplatin resistance in non-small cell lung cancer. Epigenetics 2014;9:896–909. https://doi.org/10.4161/epi.28601.
[133] Dillinger T, Sheibani-Tezerji R, Pulverer W, Stelzer I, Hassler MR, Scheibelreiter J, et al. Identification of tumor tissue-derived DNA methylation biomarkers for the detection and therapy response evaluation of metastatic castration resistant prostate cancer in liquid biopsies. Mol Cancer 2022;21:7. https://doi.org/10.1186/s12943-021-01445-0.
[134] Pedersen CA, Cao MD, Fleischer T, Rye MB, Knappskog S, Eikesdal HP, et al. DNA methylation changes in response to neoadjuvant chemotherapy are associated with breast cancer survival. Breast Cancer Res 2022;24:43. https://doi.org/10.1186/s13058-022-01537-9.
[135] Duruisseaux M, Martínez-Cardús A, Calleja-Cervantes ME, Moran S, Moura MC de, Davalos V, et al. Epigenetic prediction of response to anti-PD-1 treatment in non-small-cell lung cancer: a multicentre, retrospective analysis. Lancet Respir Med 2018;6:771–81. https://doi.org/10.1016/s2213-2600(18)30284-4.
[136] Xu B, Lu M, Yan L, Ge M, Ren Y, Wang R, et al. A Pan-Cancer Analysis of Predictive Methylation Signatures of Response to Cancer Immunotherapy. Front Immunol 2021;12:796647. https://doi.org/10.3389/fimmu.2021.796647.
[137] Frith MC, Pheasant M, Mattick JS. Genomics: The amazing complexity of the human transcriptome. Eur J Hum Genet 2005;13:894–7. https://doi.org/10.1038/sj.ejhg.5201459.
[138] Zhang P, Zucchelli M, Bruce S, Hambiliki F, Stavreus-Evers A, Levkov L, et al. Transcriptome Profiling of Human Pre-Implantation Development. PLoS ONE 2009;4:e7844. https://doi.org/10.1371/journal.pone.0007844.
[139] BUCCA G, CARRUBA G, SAETTA A, MUTI P, CASTAGNETTA L, SMITH CP. Gene Expression Profiling of Human Cancers. Ann N Y Acad Sci 2004;1028:28–37. https://doi.org/10.1196/annals.1322.003.
[140] Morin RD, Bainbridge M, Fejes A, Hirst M, Krzywinski M, Pugh TJ, et al. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. BioTechniques 2008;45:81–94. https://doi.org/10.2144/000112900.
[141] Grigoriadis A, Mackay A, Reis-Filho JS, Steele D, Iseli C, Stevenson BJ, et al. Establishment of the epithelial-specific transcriptome of normal and malignant human breast cells based on MPSS and array expression data. Breast Cancer Res 2006;8:R56. https://doi.org/10.1186/bcr1604.
[142] Li P, Conley A, Zhang H, Kim HL. Whole-Transcriptome profiling of formalin-fixed, paraffin-embedded renal cell carcinoma by RNA-seq. BMC Genom 2014;15:1087. https://doi.org/10.1186/1471-2164-15-1087.
[143] Malouf GG, Su X, Yao H, Gao J, Xiong L, He Q, et al. Next-Generation Sequencing of Translocation Renal Cell Carcinoma Reveals Novel RNA Splicing Partners and Frequent Mutations of Chromatin-Remodeling Genes. Clin Cancer Res 2014;20:4129–40. https://doi.org/10.1158/1078-0432.ccr-13-3036.
[144] Liu J, Lee W, Jiang Z, Chen Z, Jhunjhunwala S, Haverty PM, et al. Genome and transcriptome sequencing of lung cancers reveal diverse mutational and splicing events. Genome Res 2012;22:2315–27. https://doi.org/10.1101/gr.140988.112.
[145] Jones S, Zhang X, Parsons DW, Lin JC-H, Leary RJ, Angenendt P, et al. Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses. Science 2008;321:1801–6. https://doi.org/10.1126/science.1164368.
[146] Jung Y, Kim P, Jung Y, Keum J, Kim S, Choi YS, et al. Discovery of ALK-PTPN3 gene fusion from human non-small cell lung carcinoma cell line using next generation RNA sequencing. Genes, Chromosom Cancer 2012;51:590–7. https://doi.org/10.1002/gcc.21945.
[147] Berger MF, Levin JZ, Vijayendran K, Sivachenko A, Adiconis X, Maguire J, et al. Integrative analysis of the melanoma transcriptome. Genome Res 2010;20:413–27. https://doi.org/10.1101/gr.103697.109.
[148] Bridgewater J, Laar R, van, Floore A, Veer LV. Gene expression profiling may improve diagnosis in patients with carcinoma of unknown primary. Br J Cancer 2008;98:1425–30. https://doi.org/10.1038/sj.bjc.6604315.
[149] Kuruvilla FG, Park PJ, Schreiber SL. Vector algebra in the analysis of genome-wide expression data. research0011.1 Genome Biol 2002;3. https://doi.org/10.1186/gb-2002-3-3-research0011.
[150] Holter NS, Mitra M, Maritan A, Cieplak M, Banavar JR, Fedoroff NV. Fundamental patterns underlying gene expression profiles: Simplicity from complexity. Proc Natl Acad Sci 2000;97:8409–14. https://doi.org/10.1073/pnas.150242097.
[151] Yeung KY, Ruzzo WL. Principal component analysis for clustering gene expression data. Bioinformatics 2001;17:763–74. https://doi.org/10.1093/bioinformatics/17.9.763.

[152] Tothill RW, Shi F, Paiman L, Bedo J, Kowalczyk A, Mileshkin L, et al. Development and validation of a gene expression tumour classifier for cancer of unknown primary. Pathology 2015;47:7–12. https://doi.org/10.1097/pat.0000000000000194.
[153] Divate M, Tyagi A, Richard DJ, Prasad PA, Gowda H, Nagaraj SH. Deep Learning-Based Pan-Cancer Classification Model Reveals Tissue-of-Origin Specific Gene Expression Signatures. Cancers 2022;14:1185. https://doi.org/10.3390/cancers14051185.
[154] Xu Q, Chen J, Ni S, Tan C, Xu M, Dong L, et al. Pan-cancer transcriptome analysis reveals a gene expression signature for the identification of tumor tissue origin. Mod Pathol 2016;29:546–56. https://doi.org/10.1038/modpathol.2016.60.
[155] Michuda J, Breschi A, Kapilivsky J, Manghnani K, McCarter C, Hockenberry AJ, et al. Validation of a Transcriptome-Based Assay for Classifying Cancers of Unknown Primary Origin. Mol Diagn Ther 2023;27:499–511. https://doi.org/10.1007/s40291-023-00650-5.
[156] Guinney J, Dienstmann R, Wang X, Reyniès A de, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. Nat Med 2015;21:1350–6. https://doi.org/10.1038/nm.3967.
[157] Joanito I, Wirapati P, Zhao N, Nawaz Z, Yeo G, Lee F, et al. Single-cell and bulk transcriptome sequencing identifies two epithelial tumor cell states and refines the consensus molecular classification of colorectal cancer. Nat Genet 2022;54:963–75. https://doi.org/10.1038/s41588-022-01100-4.
[158] Kamoun A, Reyniès A de, Allory Y, Sjödahl G, Robertson AG, Seiler R, et al. A Consensus Molecular Classification of Muscle-invasive Bladder Cancer. Eur Urol 2020;77:420–33. https://doi.org/10.1016/j.eururo.2019.09.006.
[159] Niemira M, Collin F, Szalkowska A, Bielska A, Chwialkowska K, Reszec J, et al. Molecular Signature of Subtypes of Non-Small-Cell Lung Cancer by Large-Scale Transcriptional Profiling: Identification of Key Modules and Genes by Weighted Gene Co-Expression Network Analysis (WGCNA). Cancers 2019;12:37. https://doi.org/10.3390/cancers12010037.
[160] Monti S, Tamayo P, Mesirov J, Golub T. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. Mach Learn 2003;52:91–118. https://doi.org/10.1023/a:1023949509487.
[161] Jiang Y-Z, Ma D, Suo C, Shi J, Xue M, Hu X, et al. Genomic and Transcriptomic Landscape of Triple-Negative Breast Cancers: Subtypes and Treatment Strategies. Cancer Cell 2019;35:428–440.e5. https://doi.org/10.1016/j.ccell.2019.02.001.
[162] Cheng W-Y, Li J-F, Zhu Y-M, Lin X-J, Wen L-J, Zhang F, et al. Transcriptome-based molecular subtypes and differentiation hierarchies improve the classification framework of acute myeloid leukemia. Proc Natl Acad Sci 2022;119:e2211429119. https://doi.org/10.1073/pnas.2211429119.
[163] Li Y, John MARSt, Zhou X, Kim Y, Sinha U, Jordan RCK, et al. Salivary Transcriptome Diagnostics for Oral Cancer Detection. Clin Cancer Res 2004;10:8442–50. https://doi.org/10.1158/1078-0432.ccr-04-1167.
[164] Brinkmann O, Wong DTW. Salivary Transcriptome Biomarkers in Oral Squamous Cell Cancer Detection. Adv Clin Chem 2011;55:21–34. https://doi.org/10.1016/b978-0-12-387042-1.00002-2.
[165] Martin J.L., Gottehrer N., Zalesin H., Hoff P.T., Shaw M., Clarkson J.H., et al. Evaluation of Salivary Transcriptome Markers for the Early Detection of Oral Squamous Cell Cancer in a Prospective Blinded Trial. Compend Contin Educ Dent (Jamesburg, NJ: 1995) 2015;36:365–73.
[166] Whitney DH, Elashoff MR, Porta-Smith K, Gower AC, Vachani A, Ferguson JS, et al. Derivation of a bronchial genomic classifier for lung cancer in a prospective study of patients undergoing diagnostic bronchoscopy. BMC Méd Genom 2015;8:18. https://doi.org/10.1186/s12920-015-0091-3.
[167] Silvestri GA, Vachani A, Whitney D, Elashoff M, Smith KP, Ferguson JS, et al. A Bronchial Genomic Classifier for the Diagnostic Evaluation of Lung Cancer. N Engl J Med 2015;373:243–51. https://doi.org/10.1056/nejmoa1504601.
[168] Mazzone PJ, Lamb C, Rieger-Christ KM, Reddy CB, Qu J, Wu S, et al. Early candidate nasal swab classifiers developed using machine learning and whole transcriptome sequencing may improve early lung cancer detection. 8551–8551 J Clin Oncol 2021;39. https://doi.org/10.1200/jco.2021.39.15_suppl.8551.
[169] Ge X, Xu H, Weng S, Zhang Y, Liu L, Wang L, et al. Systematic analysis of transcriptome signature for improving outcomes in lung adenocarcinoma. J Cancer Res Clin Oncol 2023;149:8951–68. https://doi.org/10.1007/s00432-023-04814-y.
[170] Wang M, Lindberg J, Klevebring D, Nilsson C, Mer AS, Rantalainen M, et al. Validation of risk stratification models in acute myeloid leukemia using sequencing-based molecular profiling. Leukemia 2017;31:2029–36. https://doi.org/10.1038/leu.2017.48.
[171] Tao Y, Wei L, You H. Ferroptosis-related gene signature predicts the clinical outcome in pediatric acute myeloid leukemia patients and refines the 2017 ELN classification system. Front Mol Biosci 2022;9:954524. https://doi.org/10.3389/fmolb.2022.954524.
[172] Samadi P, Soleimani M, Nouri F, Rahbarizadeh F, Najafi R, Jalali A. An integrative transcriptome analysis reveals potential predictive, prognostic biomarkers and therapeutic targets in colorectal cancer. BMC Cancer 2022;22:835. https://doi.org/10.1186/s12885-022-09931-4.
[173] Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. Genome Med 2021;13:152. https://doi.org/10.1186/s13073-021-00968-x.
[174] Brown LM, Lonsdale A, Zhu A, Davidson NM, Schmidt B, Hawkins A, et al. The application of RNA sequencing for the diagnosis and genomic classification of pediatric acute lymphoblastic leukemia. Blood Adv 2020;4:930–42. https://doi.org/10.1182/bloodadvances.2019001008.
[175] Walter W, Shahswar R, Stengel A, Meggendorfer M, Kern W, Haferlach T, et al. Clinical application of whole transcriptome sequencing for the classification of patients with acute lymphoblastic leukemia. BMC Cancer 2021;21:886. https://doi.org/10.1186/s12885-021-08635-5.
[176] Docking TR, Parker JDK, Jädersten M, Duns G, Chang L, Jiang J, et al. A clinical transcriptome approach to patient stratification and therapy selection in acute myeloid leukemia. Nat Commun 2021;12:2474. https://doi.org/10.1038/s41467-021-22625-y.
[177] Backert L, Kohlbacher O. Immunoinformatics and epitope prediction in the age of genomic medicine. Genome Med 2015;7:119. https://doi.org/10.1186/s13073-015-0245-0.
[178] Peng K, Nowicki TS, Campbell K, Vahed M, Peng D, Meng Y, et al. Rigorous benchmarking of T-cell receptor repertoire profiling methods for cancer RNA sequencing. Brief Bioinform 2023;24:bbad220. https://doi.org/10.1093/bib/bbad220.
[179] Chen P-L, Roh W, Reuben A, Cooper ZA, Spencer CN, Prieto PA, et al. Analysis of Immune Signatures in Longitudinal Tumor Samples Yields Insight into Biomarkers of Response and Mechanisms of Resistance to Immune Checkpoint Blockade. Cancer Discov 2016;6:827–37. https://doi.org/10.1158/2159-8290.cd-15-1545.
[180] Macklin A, Khan S, Kislinger T. Recent advances in mass spectrometry based clinical proteomics: applications to cancer research. Clin Prote 2020;17:17. https://doi.org/10.1186/s12014-020-09283-w.
[181] Katz DH, Robbins JM, Deng S, Tahir UA, Bick AG, Pampana A, et al. Proteomic profiling platforms head to head: Leveraging genetics and clinical traits to compare aptamer- and antibody-based methods. Sci Adv 2022;8:eabm5164. https://doi.org/10.1126/sciadv.abm5164.
[182] Ku X, Cai C, Xu Y, Chen S, Zhou Z, Xiao J, et al. Data independent acquisition-mass spectrometry (DIA-MS)-based comprehensive profiling of bone metastatic cancers revealed molecular fingerprints to assist clinical classifications for bone metastasis of unknown primary (BMUP). Transl Cancer Res 2020;9:2390–401. https://doi.org/10.21037/tcr.2020.03.41.
[183] Boys E, Liu J (Jenny, Cai Z, Noor Z, Mackenzie KL, Aref A, et al. Pan-cancer diagnostic proteomic signature of tissue-of-origin (TOO) using data-independent acquisition mass spectrometry (DIA-MS) from 1289 human tissue samples. 3120–3120 J Clin Oncol 2023;41. https://doi.org/10.1200/jco.2023.41.16_suppl.3120.
[184] Ellis MJ, Gillette M, Carr SA, Paulovich AG, Smith RD, Rodland KK, et al. Connecting Genomic Alterations to Cancer Biology with Proteomics: The NCI Clinical Proteomic Tumor Analysis Consortium. Cancer Discov 2013;3:1108–12. https://doi.org/10.1158/2159-8290.cd-13-0219.
[185] Mani DR, Maynard M, Kothadia R, Krug K, Christianson KE, Heiman D, et al. PANOPLY: a cloud-based platform for automated and reproducible proteogenomic data analysis. Nat Methods 2021;18:580–2. https://doi.org/10.1038/s41592-021-01176-6.
[186] 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;166:755–65. https://doi.org/10.1016/j.cell.2016.05.069.
[187] Zhang B, Wang J, Wang X, Zhu J, Liu Q, Shi Z, et al. Proteogenomic characterization of human colon and rectal cancer. Nature 2014;513:382–7. https://doi.org/10.1038/nature13438.
[188] Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature 1999;401:788–91. https://doi.org/10.1038/44565.
[189] Jiang Y, Sun A, Zhao Y, Ying W, Sun H, Yang X, et al. Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. Nature 2019;567:257–61. https://doi.org/10.1038/s41586-019-0987-8.
[190] Krug K, Jaehnig EJ, Satpathy S, Blumenberg L, Karpova A, Anurag M, et al. Proteogenomic Landscape of Breast Cancer Tumorigenesis and Targeted Therapy. Cell 2020;183:1436–1456.e31. https://doi.org/10.1016/j.cell.2020.10.036.
[191] Cao L, Huang C, Zhou DC, Hu Y, Lih TM, Savage SR, et al. Proteogenomic characterization of pancreatic ductal adenocarcinoma. Cell 2021;184:5031–5052.e26. https://doi.org/10.1016/j.cell.2021.08.023.
[192] Wang, Karpova L-B, Gritsenko MA A, Kyle JE, Cao S, Li Y, et al. Proteogenomic and metabolomic characterization of human glioblastoma. Cancer Cell 2021;39:509–528.e20. https://doi.org/10.1016/j.ccell.2021.01.006.
[193] Pavlova NN, Zhu J, Thompson CB. The hallmarks of cancer metabolism: Still emerging. Cell Metab 2022;34:355–77. https://doi.org/10.1016/j.cmet.2022.01.007.
[194] Gal J, Bailleux C, Chardin D, Pourcher T, Gilhodes J, Jing L, et al. Comparison of unsupervised machine-learning methods to identify metabolomic signatures in patients with localized breast cancer. Comput Struct Biotechnol J 2020;18:1509–24. https://doi.org/10.1016/j.csbj.2020.05.021.
[195] Liu X, Cheng X, Liu X, He L, Zhang W, Wang Y, et al. Investigation of the urinary metabolic variations and the application in bladder cancer biomarker discovery. Int J Cancer 2018;143:408–18. https://doi.org/10.1002/ijc.31323.
[196] Tan G, Wang H, Yuan J, Qin W, Dong X, Wu H, et al. Three serum metabolite signatures for diagnosing low-grade and high-grade bladder cancer. Sci Rep 2017;7:46176. https://doi.org/10.1038/srep46176.
[197] Meo, di NA, Loizzo D, Pandolfo SD, Autorino R, Ferro M, Porta C, et al. Metabolomic Approaches for Detection and Identification of Biomarkers and Altered Pathways in Bladder Cancer. Int J Mol Sci 2022;23:4173. https://doi.org/10.3390/ijms23084173.
[198] Subramani R, Poudel S, Smith KD, Estrada A, Lakshmanaswamy R. Metabolomics of Breast Cancer: A Review. Metabolites 2022;12:643. https://doi.org/10.3390/metabo12070643.

[199] Wang X, Zhao X, Zhao J, Yang T, Zhang F, Liu L. Serum metabolite signatures of epithelial ovarian cancer based on targeted metabolomics. Clin Chim Acta 2021; 518:59–69. https://doi.org/10.1016/j.cca.2021.03.012.
[200] Ananieva EA, Wilkinson AC. Branched-chain amino acid metabolism in cancer. Curr Opin Clin Nutr Metab Care 2018;21:64–70. https://doi.org/10.1097/mco.0000000000000430.
[201] Wulfkuhle JD, Liotta LA, Petricoin EF. Proteomic applications for the early detection of cancer. Nat Rev Cancer 2003;3:267–75. https://doi.org/10.1038/nrc1043.
[202] Adam B-L, Qu Y, Davis JW, Ward MD, Clements MA, Cazares LH, et al. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. Cancer Res 2002;62:3609–14.
[203] Petricoin EF, Ornstein DK, Paweletz CP, Ardekani A, Hackett PS, Hitt BA, et al. Serum Proteomic Patterns for Detection of Prostate Cancer. JNCI: J Natl Cancer Inst 2002;94:1576–8. https://doi.org/10.1093/jnci/94.20.1576.
[204] Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, et al. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 2002;359: 572–7. https://doi.org/10.1016/s0140-6736(02)07746-2.
[205] Kohonen T. The self-organizing map. Proc IEEE 1990;78:1464–80. https://doi.org/10.1109/5.58325.
[206] Belluco C, Petricoin EF, Mammano E, Facchiano F, Ross-Rucker S, Nitti D, et al. Serum Proteomic Analysis Identifies a Highly Sensitive and Specific Discriminatory Pattern in Stage 1 Breast Cancer. Ann Surg Oncol 2007;14: 2470–6. https://doi.org/10.1245/s10434-007-9354-3.
[207] Li B, Kugeratski FG, Kalluri R. A novel machine learning algorithm selects proteome signature to specifically identify cancer exosomes. 2023.07.18.549557 BioRxiv 2023. https://doi.org/10.1101/2023.07.18.549557.
[208] Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018;562: 203–9. https://doi.org/10.1038/s41586-018-0579-z.
[209] Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature 2023;613:508–18. https://doi.org/10.1038/s41586-022-05473-8.
[210] Sanderson E, Glymour MM, Holmes MV, Kang H, Morrison J, Munafò MR, et al. Mendelian randomization. Nat Rev Methods Prim 2022;2:6. https://doi.org/10.1038/s43586-021-00092-5.
[211] Yao C, Chen G, Song C, Keefe J, Mendelson M, Huan T, et al. Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. Nat Commun 2018;9:3268. https://doi.org/10.1038/s41467-018-05512-x.
[212] Pietzner M, Wheeler E, Carrasco-Zanini J, Kerrison ND, Oerton E, Koprulu M, et al. Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. Nat Commun 2021;12:6822. https://doi.org/10.1038/s41467-021-27164-0.
[213] Eldjarn GH, Ferkingstad E, Lund SH, Helgason H, Magnusson OTh, Gunnarsdottir K, et al. Large-scale plasma proteomics comparisons through genetics and disease associations. Nature 2023;622:348–58. https://doi.org/10.1038/s41586-023-06563-x.
[214] Mälarstig A, Grassmann F, Dahl L, Dimitriou M, McLeod D, Gabrielson M, et al. Evaluation of circulating plasma proteins in breast cancer using Mendelian randomisation. Nat Commun 2023;14:7680. https://doi.org/10.1038/s41467-023-43485-8.
[215] Sun J, Zhao J, Jiang F, Wang L, Xiao Q, Han F, et al. Identification of novel protein biomarkers and drug targets for colorectal cancer by integrating human plasma proteome with genome. Genome Med 2023;15:75. https://doi.org/10.1186/s13073-023-01229-9.
[216] Bowden J, Smith GD, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol 2015;44:512–25. https://doi.org/10.1093/ije/dyv080.
[217] Bowden J, Smith GD, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol 2016;40:304–14. https://doi.org/10.1002/gepi.21965.
[218] Verbanck M, Chen C-Y, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet 2018;50:693–8. https://doi.org/10.1038/s41588-018-0099-7.
[219] Burgess S, Foley CN, Allara E, Staley JR, Howson JMM. A robust and efficient method for Mendelian randomization with hundreds of genetic variants. Nat Commun 2020;11:376. https://doi.org/10.1038/s41467-019-14156-4.
[220] Feng Y, Wang R, Li C, Cai X, Huo Z, Liu Z, et al. Causal effects of genetically determined metabolites on cancers included lung, breast, ovarian cancer, and glioma: a Mendelian randomization study. Transl Lung Cancer Res 2022;11: 1302–14. https://doi.org/10.21037/tlcr-22-34.
[221] Asleh K, Negri GL, Miko SES, Colborne S, Hughes CS, Wang XQ, et al. Proteomic analysis of archival breast cancer clinical specimens identifies biological subtypes with distinct survival outcomes. Nat Commun 2022;13:896. https://doi.org/10.1038/s41467-022-28524-0.
[222] Soltis AR, Bateman NW, Liu J, Nguyen T, Franks TJ, Zhang X, et al. Proteogenomic analysis of lung adenocarcinoma reveals tumor heterogeneity, survival determinants, and therapeutically relevant pathways. Cell Rep Med 2022;3:100819. https://doi.org/10.1016/j.xcrm.2022.100819.
[223] Armitage EG, Southam AD. Monitoring cancer prognosis, diagnosis and treatment efficacy using metabolomics and lipidomics. Metabolomics 2016;12:146. https://doi.org/10.1007/s11306-016-1093-7.
[224] Wang W, Rong Z, Wang G, Hou Y, Yang F, Qiu M. Cancer metabolites: promising biomarkers for cancer liquid biopsy. Biomark Res 2023;11:66. https://doi.org/10.1186/s40364-023-00507-3.
[225] Zhu J, Djukovic D, Deng L, Gu H, Himmati F, Zaid MA, et al. Targeted serum metabolite profiling and sequential metabolite ratio analysis for colorectal cancer progression monitoring. Anal Bioanal Chem 2015;407:7857–63. https://doi.org/10.1007/s00216-015-8984-8.
[226] Lodi A, Tiziani S, Khanim FL, Günther UL, Viant MR, Morgan GJ, et al. Proton NMR-Based Metabolite Analyses of Archived Serial Paired Serum and Urine Samples from Myeloma Patients at Different Stages of Disease Activity Identifies Acetylcarnitine as a Novel Marker of Active Disease. PLoS ONE 2013;8:e56422. https://doi.org/10.1371/journal.pone.0056422.
[227] Xing X, Hu E, Ouyang J, Zhong X, Wang F, Liu K, et al. Integrated omics landscape of hepatocellular carcinoma suggests proteomic subtypes for precision therapy. Cell Rep Med 2023;4:101315. https://doi.org/10.1016/j.xcrm.2023.101315.
[228] Kim H, Yu SJ, Yeo I, Cho YY, Lee DH, Cho Y, et al. Prediction of Response to Sorafenib in Hepatocellular Carcinoma: A Putative Marker Panel by Multiple Reaction Monitoring-Mass Spectrometry (MRM-MS)*. Mol Cell Prote 2017;16: 1312–23. https://doi.org/10.1074/mcp.m116.066704.
[229] Li Y, Xu C, Wang B, Xu F, Ma F, Qu Y, et al. Proteomic characterization of gastric cancer response to chemotherapy and targeted therapy reveals potential therapeutic strategies. Nat Commun 2022;13:5723. https://doi.org/10.1038/s41467-022-33282-0.
[230] Cordo' V, Meijer MT, Hagelaar R, Haas RR de G, Poort VM, Henneman AA, et al. Phosphoproteomic profiling of T cell acute lymphoblastic leukemia reveals targetable kinases and combination treatment strategies. Nat Commun 2022;13: 1048. https://doi.org/10.1038/s41467-022-28682-1.
[231] Taguchi F, Solomon B, Gregorc V, Roder H, Gray R, Kasahara K, et al. Mass Spectrometry to Classify Non–Small-Cell Lung Cancer Patients for Clinical Outcome After Treatment With Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitors: A Multicohort Cross-Institutional Study. JNCI: J Natl Cancer Inst 2007;99:838–46. https://doi.org/10.1093/jnci/djk195.
[232] Carbone DP, Ding K, Roder H, Grigorieva J, Roder J, Tsao M-S, et al. Prognostic and Predictive Role of the VeriStrat Plasma Test in Patients with Advanced Non–Small-Cell Lung Cancer Treated with Erlotinib or Placebo in the NCIC Clinical Trials Group BR.21 Trial. J Thorac Oncol 2012;7:1653–60. https://doi.org/10.1097/jto.0b013e31826c1155.
[233] Gregorc V, Novello S, Lazzari C, Barni S, Aieta M, Mencoboni M, et al. Predictive value of a proteomic signature in patients with non-small-cell lung cancer treated with second-line erlotinib or chemotherapy (PROSE): a biomarker-stratified, randomised phase 3 trial. Lancet Oncol 2014;15:713–21. https://doi.org/10.1016/s1470-2045(14)70162-7.
[234] Fidler MJ, Fhied CL, Roder J, Basu S, Sayidine S, Fughhi I, et al. The serum-based VeriStrat® test is associated with proinflammatory reactants and clinical outcome in non-small cell lung cancer patients. BMC Cancer 2018;18:310. https://doi.org/10.1186/s12885-018-4193-0.
[235] Chae YK, Kim WB, Davis AA, Park LC, Anker JF, Simon NI, et al. Mass spectrometry-based serum proteomic signature as a potential biomarker for survival in patients with non-small cell lung cancer receiving immunotherapy. Transl Lung Cancer Res 2020;9:1015–28. https://doi.org/10.21037/tlcr-20-148.
[236] Rich P, Mitchell RB, Schaefer E, Walker PR, Dubay JW, Boyd J, et al. Real-world performance of blood-based proteomic profiling in first-line immunotherapy treatment in advanced stage non-small cell lung cancer. J Immunother Cancer 2021;9:e002989. https://doi.org/10.1136/jitc-2021-002989.
[237] Harel M, Ortenberg R, Varanasi SK, Mangalhara KC, Mardamshina M, Markovits E, et al. Proteomics of Melanoma Response to Immunotherapy Reveals Mitochondrial Dependence. e18 Cell 2019;179:236–50. https://doi.org/10.1016/j.cell.2019.08.012.
[238] Muller M, Hummelink K, Hurkmans DP, Niemeijer A-LN, Monkhorst K, Roder J, et al. A Serum Protein Classifier Identifying Patients with Advanced Non–Small Cell Lung Cancer Who Derive Clinical Benefit from Treatment with Immune Checkpoint Inhibitors. Clin Cancer Res 2020;26:5188–97. https://doi.org/10.1158/1078-0432.ccr-20-0538.
[239] Wang J, Sun N, Kunzke T, Shen J, Zens P, Prade VM, et al. Spatial metabolomics identifies distinct tumor-specific and stroma-specific subtypes in patients with lung squamous cell carcinoma. Npj Precis Oncol 2023;7:114. https://doi.org/10.1038/s41698-023-00434-4.
[240] Jurmeister P, Bockmayr M, Seegerer P, Bockmayr T, Treue D, Montavon G, et al. Machine learning analysis of DNA methylation profiles distinguishes primary lung squamous cell carcinomas from head and neck metastases. Sci Transl Med 2019; 11. https://doi.org/10.1126/scitranslmed.aaw8513.
[241] Klughammer J, Kiesel B, Roetzer T, Fortelny N, Nemc A, Nenning K-H, et al. The DNA methylation landscape of glioblastoma disease progression shows extensive heterogeneity in time and space. Nat Med 2018;24:1611–24. https://doi.org/10.1038/s41591-018-0156-x.
[242] Rusch M, Nakitandwe J, Shurtleff S, Newman S, Zhang Z, Edmonson MN, et al. Clinical cancer genomic profiling by three-platform sequencing of whole genome, whole exome and transcriptome. Nat Commun 2018;9:3962. https://doi.org/10.1038/s41467-018-06485-7.
[243] Leongamornlert D, Gutiérrez-Abril J, Lee SW, Barretta E, Creasey T, Gundem G, et al. Diagnostic utility of whole genome sequencing in adults with B-other acute lymphoblastic leukemia. Blood Adv 2023;7:3862–73. https://doi.org/10.1182/bloodadvances.2022008992.
[244] Belzen IAEM van, Cai C, Tuil M, van, Badloe S, Strengman E, Janse A, et al. Systematic discovery of gene fusions in pediatric cancer by integrating RNA-seq

and WGS. BMC Cancer 2023;23:618. https://doi.org/10.1186/s12885-023-11054-3.
[245] Shukla N, Levine MF, Gundem G, Domenico D, Spitzer B, Bouvier N, et al. Feasibility of whole genome and transcriptome profiling in pediatric and young adult cancers. Nat Commun 2022;13:2485. https://doi.org/10.1038/s41467-022-30233-7.
[246] Pleasance E, Bohm A, Williamson LM, Nelson JMT, Shen Y, Bonakdar M, et al. Whole-genome and transcriptome analysis enhances precision cancer treatment options. Ann Oncol 2022;33:939–49. https://doi.org/10.1016/j.annonc.2022.05.522.
[247] Wong M, Mayoh C, Lau LMS, Khuong-Quang D-A, Pinese M, Kumar A, et al. Whole genome, transcriptome and methylome profiling enhances actionable target discovery in high-risk pediatric cancer. Nat Med 2020;26:1742–53. https://doi.org/10.1038/s41591-020-1072-4.
[248] Xu J, Wu P, Chen Y, Meng Q, Dawood H, Dawood H. A hierarchical integration deep flexible neural forest framework for cancer subtype classification by integrating multi-omics data. BMC Bioinform 2019;20:527. https://doi.org/10.1186/s12859-019-3116-7.
[249] Ren B, Yang J, Wang C, Yang G, Wang H, Chen Y, et al. High-resolution Hi-C maps highlight multiscale 3D epigenome reprogramming during pancreatic cancer metastasis. J Hematol Oncol 2021;14:120. https://doi.org/10.1186/s13045-021-01131-0.
[250] Sanghi A, Gruber JJ, Metwally A, Jiang L, Reynolds W, Sunwoo J, et al. Chromatin accessibility associates with protein-RNA correlation in human cancer. Nat Commun 2021;12:5732. https://doi.org/10.1038/s41467-021-25872-1.
[251] Huang D, Ma N, Li X, Gou Y, Duan Y, Liu B, et al. Advances in single-cell RNA sequencing and its applications in cancer research. J Hematol Oncol 2023;16:98. https://doi.org/10.1186/s13045-023-01494-6.
[252] Aissa AF, Islam ABMMK, Ariss MM, Go CC, Rader AE, Conrardy RD, et al. Single-cell transcriptional changes associated with drug tolerance and response to combination therapies in cancer. Nat Commun 2021;12:1628. https://doi.org/10.1038/s41467-021-21884-z.
[253] Li K, Du Y, Cai Y, Liu W, Lv Y, Huang B, et al. Single-cell analysis reveals the chemotherapy-induced cellular reprogramming and novel therapeutic targets in relapsed/refractory acute myeloid leukemia. Leukemia 2023;37:308–25. https://doi.org/10.1038/s41375-022-01789-6.
[254] Lim J, Chin V, Fairfax K, Moutinho C, Suan D, Ji H, et al. Transitioning single-cell genomics into the clinic. Nat Rev Genet 2023;24:573–84. https://doi.org/10.1038/s41576-023-00613-w.
[255] Shouval R, Shlush LI, Yehudai-Resheff S, Ali S, Pery N, Shapiro E, et al. Single cell analysis exposes intratumor heterogeneity and suggests that FLT3-ITD is a late event in leukemogenesis. Exp Hematol 2014;42:457–63. https://doi.org/10.1016/j.exphem.2014.01.010.
[256] Robinson TM, Bowman RL, Persaud S, Liu Y, Neigenfind R, Gao Q, et al. Single-cell genotypic and phenotypic analysis of measurable residual disease in acute myeloid leukemia. Sci Adv 2023;9:eadg0488. https://doi.org/10.1126/sciadv.adg0488.
[257] Picard M, Scott-Boyer M-P, Bodein A, Périn O, Droit A. Integration strategies of multi-omics data for machine learning analysis. Comput Struct Biotechnol J 2021;19:3735–46. https://doi.org/10.1016/j.csbj.2021.06.030.
[258] Sammut S-J, Crispin-Ortuzar M, Chin S-F, Provenzano E, Bardwell HA, Ma W, et al. Multi-omic machine learning predictor of breast cancer therapy response. Nature 2022;601:623–9. https://doi.org/10.1038/s41586-021-04278-5.
[259] Islam MdM, Huang S, Ajwad R, Chi C, Wang Y, Hu P. An integrative deep learning framework for classifying molecular subtypes of breast cancer. Comput Struct Biotechnol J 2020;18:2185–99. https://doi.org/10.1016/j.csbj.2020.08.005.
[260] Foersch S, Glasner C, Woerl A-C, Eckstein M, Wagner D-C, Schulz S, et al. Multistain deep learning for prediction of prognosis and therapy response in colorectal cancer. Nat Med 2023;29:430–9. https://doi.org/10.1038/s41591-022-02134-1.
[261] Schmidt B, Brown LM, Ryland GL, Lonsdale A, Kosasih HJ, Ludlow LE, et al. ALLSorts: an RNA-Seq subtype classifier for B-cell acute lymphoblastic leukemia. Blood Adv 2022;6:4093–7. https://doi.org/10.1182/bloodadvances.2021005894.
[262] Metzeler KH, Hummel M, Bloomfield CD, Spiekermann K, Braess J, Sauerland M-C, et al. An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. Blood 2008;112:4193–201. https://doi.org/10.1182/blood-2008-02-134411.
[263] Cheng X, Liu X, Liu X, Guo Z, Sun H, Zhang M, et al. Metabolomics of non-muscle invasive bladder cancer: biomarkers for early detection of bladder cancer. Front Oncol 2018;8:494. https://doi.org/10.3389/fonc.2018.00494.
[264] Zheng S, Cherniack AD, Dewal N, Moffitt RA, Danilova L, Murray BA, et al. Comprehensive Pan-Genomic Characterization of Adrenocortical Carcinoma. Cancer Cell 2016;29:723–36. https://doi.org/10.1016/j.ccell.2016.04.002.
[265] Liu J (Jenny), Jackson C, Anees A, Noor Z, Williams S, Porceddu S, et al. A proteomic signature associated with prognosis in HPV-related locally advanced oropharyngeal squamous cell carcinoma (LA-OPSCC). 6055–6055 J Clin Oncol 2023;41. https://doi.org/10.1200/jco.2023.41.16_suppl.6055.
[266] Chang Y, Park H, Yang H-J, Lee S, Lee K-Y, Kim TS, et al. Cancer Drug Response Profile scan (CDRscan): a deep learning model that predicts drug effectiveness from cancer genomic signature. Sci Rep 2018;8:8857. https://doi.org/10.1038/s41598-018-27214-6.
[267] Grewal JK, Tessier-Cloutier B, Jones M, Gakkhar S, Ma Y, Moore R, et al. Application of a neural network whole transcriptome–based pan-cancer method for diagnosis of primary and metastatic cancers. JAMA Netw Open 2019;2:e192597. https://doi.org/10.1001/jamanetworkopen.2019.2597.
[268] Haradhvala NJ, Kim J, Maruvka YE, Polak P, Rosebrock D, Livitz D, et al. Distinct mutational signatures characterize concurrent loss of polymerase proofreading and mismatch repair. Nat Commun 2018;9:1746. https://doi.org/10.1038/s41467-018-04002-4.
[269] Prosz A, Duan H, Tisza V, Sahgal P, Topka S, Klus GT, et al. Nucleotide excision repair deficiency is a targetable therapeutic vulnerability in clear cell renal cell carcinoma. 2023.02.07.527498 BioRxiv 2023. https://doi.org/10.1101/2023.02.07.527498.
[270] Póti Á, Gyergyák H, Németh E, Rusz O, Tóth S, Kovácsházi C, et al. Correlation of homologous recombination deficiency induced mutational signatures with sensitivity to PARP inhibitors and cytotoxic agents. Genome Biol 2019;20:240. https://doi.org/10.1186/s13059-019-1867-0.
[271] Golan T, O'Kane GM, Denroche RE, Raitses-Gurevich M, Grant RC, Holter S, et al. Genomic features and classification of homologous recombination deficient pancreatic ductal adenocarcinoma. Gastroenterology 2021;160:2119–2132.e9. https://doi.org/10.1053/j.gastro.2021.01.220.
[272] Tsang ES, Csizmok V, Williamson LM, Pleasance E, Topham JT, Karasinska JM, et al. Homologous recombination deficiency signatures in gastrointestinal and thoracic cancers correlate with platinum therapy duration. Npj Precis Oncol 2023;7:31. https://doi.org/10.1038/s41698-023-00368-x.
[273] Yazinski SA, Comaills V, Buisson R, Genois M-M, Nguyen HD, Ho CK, et al. ATR inhibition disrupts rewired homologous recombination and fork protection pathways in PARP inhibitor-resistant BRCA-deficient cancer cells. Genes Dev 2017;31:318–32. https://doi.org/10.1101/gad.290957.116.
[274] Chen X, Gole J, Gore A, He Q, Lu M, Min J, et al. Non-invasive early detection of cancer four years before conventional diagnosis using a blood test. Nat Commun 2020;11:3475. https://doi.org/10.1038/s41467-020-17316-z.
[275] Hao X, Luo H, Krawczyk M, Wei W, Wang W, Wang J, et al. DNA methylation markers for diagnosis and prognosis of common cancers. Proc Natl Acad Sci 2017;114:7414–9. https://doi.org/10.1073/pnas.1703577114.
[276] Decock A, Ongenaert M, Cannoodt R, Verniers K, Wilde BD, Laureys G, et al. Methyl-CpG-binding domain sequencing reveals a prognostic methylation signature in neuroblastoma. Oncotarget 2015;7:1960–72. https://doi.org/10.18632/oncotarget.6477.
[277] Luo R, Song J, Xiao X, Xie Z, Zhao Z, Zhang W, et al. Identifying CpG methylation signature as a promising biomarker for recurrence and immunotherapy in non–small-cell lung carcinoma. Aging (Albany NY) 2020;12:14649–76. https://doi.org/10.18632/aging.103517.
[278] Ressler JM, Tomasich E, Hatziioannou T, Ringl H, Heller G, Silmbrod R, et al. Correlation of DNA methylation signatures with response to immune checkpoint inhibitors in metastatic melanoma. 9561–9561 J Clin Oncol 2023;41. https://doi.org/10.1200/jco.2023.41.16_suppl.9561.
[279] Horr C, Buechler SA. Breast Cancer Consensus Subtypes: a system for subtyping breast cancer tumors based on gene expression. Npj Breast Cancer 2021;7:136. https://doi.org/10.1038/s41523-021-00345-2.
[280] Chereda H, Bleckmann A, Menck K, Perera-Bel J, Stegmaier P, Auer F, et al. Explaining decisions of graph convolutional neural networks: patient-specific molecular subnetworks responsible for metastasis prediction in breast cancer. Genome Med 2021;13:42. https://doi.org/10.1186/s13073-021-00845-7.
[281] Du K, Wei S, Wei Z, Frederick DT, Miao B, Moll T, et al. Pathway signatures derived from on-treatment tumor specimens predict response to anti-PD1 blockade in metastatic melanoma. Nat Commun 2021;12:6023. https://doi.org/10.1038/s41467-021-26299-4.
[282] Wang L, Luo X, Cheng C, Amos CI, Cai G, Xiao F. A gene expression-based immune signature for lung adenocarcinoma prognosis. Cancer Immunol, Immunother 2020;69:1881–90. https://doi.org/10.1007/s00262-020-02595-8.