Multi-omics analysis: Paving the path toward achieving precision medicine in cancer treatment and immuno-oncology

Virgile Raufaste-Cazavieille[1], Raoul Santiago[1,2]*† and Arnaud Droit[1]*†

[1]CHU de Québec Research Center, Université Laval, Québec, QC, Canada, [2]Division of Pediatric Hematology-Oncology, Centre Hospitalier Universitaire de L'Université Laval, Charles Bruneau Cancer Center, Québec, QC, Canada

The acceleration of large-scale sequencing and the progress in high-throughput computational analyses, defined as omics, was a hallmark for the comprehension of the biological processes in human health and diseases. In cancerology, the omics approach, initiated by genomics and transcriptomics studies, has revealed an incredible complexity with unsuspected molecular diversity within a same tumor type as well as spatial and temporal heterogeneity of tumors. The integration of multiple biological layers of omics studies brought oncology to a new paradigm, from tumor site classification to pan-cancer molecular classification, offering new therapeutic opportunities for precision medicine. In this review, we will provide a comprehensive overview of the latest innovations for multi-omics integration in oncology and summarize the largest multi-omics dataset available for adult and pediatric cancers. We will present multi-omics techniques for characterizing cancer biology and show how multi-omics data can be combined with clinical data for the identification of prognostic and treatment-specific biomarkers, opening the way to personalized therapy. To conclude, we will detail the newest strategies for dissecting the tumor immune environment and host−tumor interaction. We will explore the advances in immunomics and microbiomics for biomarker identification to guide therapeutic decision in immuno-oncology.

KEYWORDS

multi-omics, machine learning, immunology, immunomics, microbiome, cancer, precision medicine

Introduction

Cancer is an important cause of death worldwide, even if its mortality has declined during the last decade (Santucci et al., 2020). The International Agency for Research on Cancer (IARC) GLOBOCAN cancer statistics predicted an increase of 50% for the cancer incidence and an increase of 62.5% of the mortality from now to 2040 worldwide (Ferlay et al., 2020). There is an urgent need to better cancer survival rate; however, to cure this disease, we first need to understand its underlying mechanisms.

Until recently, cancer was widely considered as organ dependent and characterized by the site of apparition of the tumor. This hypothesis was slowly abandoned due to the heterogeneity present between two patients with a similar tumor (Dagogo-Jack and Shaw, 2018) that was identified through the emergence and democratization of next-generation sequencing (NGS). NGS allows the generation of a large amount of data in a short period of time, marking the beginning of a new era for cancer research: the genomics era (Knox, 2010; Lakshmanan et al., 2020). This led to a new characterization of cancer, not based on the tumor site but founded on molecular classification (pan-classification) (Hoadley et al., 2014; Campbell et al., 2018). Following the expansion of genomics, the exploration of the other layers of cancer biology started. The apparition of different omics such as transcriptomics, epigenomics, and proteomics allowed the possibility to understand the underlying complexity of tumors (Alyass et al., 2015). The dawn of omics studies led to the discovery of further complexity with cancer presenting intra-tumoral heterogeneity at a cellular level (Dagogo-Jack and Shaw, 2018). The study of these new omics highlighted an unsuspected complexity inside the tumor architecture but also furthered our understanding of interactions between cancer and its environment (gene-environment, microenvironmental interaction, and immune system interaction) (McAllister et al., 2017; Zhou et al., 2017; Gonzalez et al., 2018; Barriga et al., 2019).

The use of single omics, such as genomics and transcriptomics, uncovered many driver genes to better comprehend the genomic landscape of cancer. Interestingly, some of these studies revealed the wide complementarity of genomics and transcriptomics, many of the driver genes being identified by either one or the other modality (Wong et al., 2020; Berlanga et al., 2022). This fosters the need of combining the interconnected biological elements of cancer to move from single-omics to multi-omics analysis. The multi-omics integration is defined by the modelization of more than one biological element in order to characterize biological systems in its globality at the phenomenological level (de Anda-Jáuregui and Hernández-Lemus, 2020). The purpose of doing this is to look at how the different biological layers of the cells interact with each other, leading to the creation of an interconnected network highlighting the underlying complexity of cancer. Data integration in cancer have three main goals: understanding the molecular mechanism of cancer, clustering disease samples, and predicting an outcome (survival or therapy efficacy) (Hiley et al., 2014; Jamal-Hanjani et al., 2014; Tebani et al., 2016; Sharifi-Noghabi et al., 2019). Computational methods were needed to integrate the large diversity of data prompting the development of new algorithms to overcome the intricacies of multi-omics integration.

To this day, even with the new information that multi-omics approaches can bring, understanding how cancer develops and maintains is puzzling. Indeed, cancer cells interact with many different components including the host immune system (Witkowski et al., 2020). These interactions define the tumor immune microenvironment (TiME) that can be beneficial or detrimental to cancer cells, leaning toward more tolerogenicity or immunogenicity (Iwai et al., 2002; Garrido and Aptsiauri, 2019; Gou et al., 2020; Tang et al., 2020). To understand the TiME provides insights on how the cancer cells hijack the immune system to survive, but also might predict if a tumor is likely to respond to immunotherapy by immune checkpoint blockade (ICB) (Riaz et al., 2017). Indeed, biomarkers derived from the study of TiME appeared to be helpful to anticipate cancer sensitivity to immunotherapy (Goswami et al., 2020). Immunomics, the field of omics-based analysis that aims to describe the reaction of the immune system to another biological component (pathogen or cancer), can be used to analyze the interactions existing between the host immune system and the cancer cells and how it leads to immune recognition or immune ignorance (Arnaout et al., 2021). However, to fully depict the interconnection that exists between the tumor and the immune system, other players need to be taken into the equation. Indeed, it has been shown that the microbiome can influence the sensitivity of cancers to ICB, suggesting an interplay between the host microbiota and the immune constitution of tumor microenvironment (Baruch et al., 2021). Immunomics and microbiomics are two novel components that should be included into multi-omics models to integrate the tumor/host interaction in cancer complexity.

In the first part, due to the growing interest in multi-omics, this review will address the challenges raised by omics and how to overcome them. In the second part, we will provide an overview of different databases that are useful in cancer research. In the third part, this review will attempt to illustrate how to integrate multi-omics to clarify cancer complexity, and how the use of machine learning can help to predict survival and treatment response. Finally, we will give an overview of new methods to decipher the interaction between the host and the cancer cells and how it can bring opportunity for personalized therapy (Figure 1).

Challenges

To integrate the massive data flow generated through the different biological elements explored by NGS, innovative computational tools are needed for diminishing the data dimension, then making them manipulable by human hands. These tools allow researchers to fulfill the gap of missing knowledges and contribute to the discovery of novel biomarkers, deciphering the complexity of cancers (Subramanian et al., 2020; Poirion et al., 2021; Wörheide et al., 2021; Zeng et al., 2021). Despite the progresses in these new instruments, multiple challenges remain.
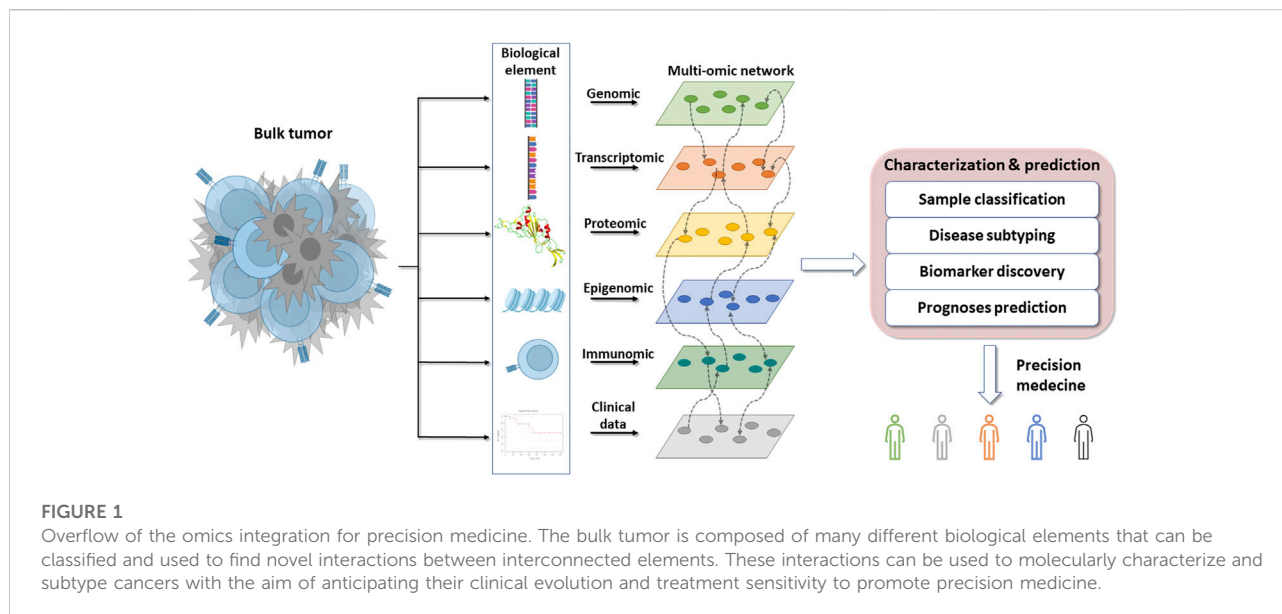
FIGURE 1
Overflow of the omics integration for precision medicine. The bulk tumor is composed of many different biological elements that can be classified and used to find novel interactions between interconnected elements. These interactions can be used to molecularly characterize and subtype cancers with the aim of anticipating their clinical evolution and treatment sensitivity to promote precision medicine.

TABLE 1 Description of different tools for multi-omics integration with their application and their major strength and limits.

Method	Principle	Aim	Omics element	Pros	Cons
JIVE	Matrix factorization	Disease subtyping, systemic knowledge, module detection	Genomics and epigenomics	Integrate large amount of data	Sensitive to outliers and missing values
NMF		Disease subtyping, module detection, biomarker discovery	Genomics and epigenomics	Filtering weak signal. Integrate large amount of data. Detection of cluster of small size	Time and memory consuming. Underperforming on missing values
nNMF					
jNMF					
intNMF					
SLIDE		Disease subtyping, module detection, biomarker discovery	Genomics, epigenomics and proteomics	Integrate large amount of data	Underperforming with missing values. Optimum solution is not guaranteed
MALA	Logic data mining	Sample classification	Genomics and transcriptomics	Works well on experimental data. Integrate large amount of data	Phenotype number must be delivered with data. Sensitive to missing values
iCluster	Gaussian latent variable model		Genomics, epigenomics and transcriptomics		Needs to test a large amount of solution to find the most relevant
iCluster+	Generalized linear regression	Disease subtyping	Genomics, transcriptomics, proteomics and epigenomics	Handle missing values	No evaluation of statistical significance for selected features
iClusterBayes	Bayesian integrative clustering	Biomarker discovery	Genomics, transcriptomics, and epigenomics	Good performance in the presence of explicative data	Underperform with outliers
MOFA	Bayesian factor analysis	Biomarker discovery, systemic knowledge	Proteomics, metabolomics and lipidomics	Handle well missing values	Linear model can miss linear relation
MOFA+			Genomics and epigenomics	The use of continuous learning enabling MOFA to recover different trajectory	Need of multi-modal measurement for the same set of cells

JIVE: joint and individual variation explained.
(n, j, int) NMF: (network, joint, integrative) non-negative matrix factorization.
SLIDE: structural learning and integrative decomposition.
MALA: micro array logic analyzer.
MOFA: multi-omics factor analysis.

One of these challenges is the heterogeneity that exists across the biological layers. It may differ in type, with numerical or categorical features, discrete or continuous variables with different ranges. The number of features can also vary between the different data sources, creating novel struggles to consider. Another burden, operating during sample collection, are missing values, setting up uninterpretable elements difficult to handle by some algorithms. Also, processing outliers and highly correlated variables might be burdensome to integrate as they might be irrelevant, create noise, or overfeed a system (Mirza et al., 2019; Song et al., 2020a).

To face these challenges, different tools have been created that we will summarize (Canzler et al., 2020; Nicora et al., 2020; Subramanian et al., 2020). Two main types of algorithms can be distinguished: 1) the "exploratory matrix" and 2) the "probabilistic matrix". The first one is based on the matrix and its result. It performs best with a well-defined matrix, can handle outliers, but is less accurate when data are missing (Devarajan, 2008; Chauvel et al., 2020; Hamamoto et al., 2022). The probabilistic matrix utilizes probabilistic formulas (such as Gaussian and Bayesian). These algorithms shine when the dataset is incomplete, with missing values, but lose accuracy with outliers (Needham et al., 2007; Yuan et al., 2021; Chu et al., 2022). This approach aims to reduce the size of the matrix to identify patterns within the dataset, allowing for powerful classification models (Table 1).

Exploratory matrix

Starting with exploratory matrices, herein is a brief overview of two algorithms using this approach: matrix factorization and logic data mining. These methods would be preferred for well-defined exhaustive datasets containing extreme values that need to be considered.

Matrix factorization: joint and individual variation explained (JIVE), non-negative factorization (NMF), and structural learning and integrative decomposition (SLIDE) are some of the tools based on this approach. This algorithm uses a system of multiplication between columns and rows on a dataset decomposed in multiple matrices of smaller dimension. Each matrices have n columns, referring to n common objects, that will be multiplied by m rows, referring to other common objects. The matrix can contain different biological elements, that is, gene expression and miRNA measurement. The strength of matrix factorization is to cluster different matrices together but requires a dataset with scarce missing values (Brunet et al., 2004; Devarajan, 2008; Lock et al., 2013; Pierre-Jean et al., 2022).

Logic data mining is a multistep procedure: 1) feature binarization, that assigns a threshold to convert each feature in binary values (0 or 1); 2) feature selection, that uses a machine learning approach to select a subset of data with relevant features introduced in a model; 3) extraction of logic formulas to build the final classification model. As an example, microarray logic analyzer (MALA) is based on this approach and has been used to identify overexpressed interrelated genes and proteins involved in a common pathway (Bertolazzi et al., 2008; Wang and éditeur, 2009; Weitschek et al., 2012).

Probabilistic matrix

We will now detail four types of algorithms to better conceptualize the probabilistic matrix approach. These models follow the principles of Gaussian probability, linear regression, or Bayesian statistics and are very useful to troubleshoot datasets with missing values.

Gaussian latent variable model uses a probabilistic matrix starting with the dimension reduction of the dataset. The matrix composed of N columns and D rows will be decreased to a matrix with a lower dimensionality on D. D will be reduced in Q most relevant data. Relevant data will then be extracted to identify a pattern explaining the results. Because the algorithm is based on Gaussian probability, dataset with a Gaussian distribution perform better with this instrument than expression profiles near 0. Lower outliers close to the null value may cause misinterpretation during the analysis. However, missing values are well managed by the algorithm. A tool using this algorithm, iCluster, will be further explained in the next chapter (Li and Chen, 2016).

Generalized linear regression is one of the subcategories found in the generalized linear model. This algorithm is composed of three components: 1) random component; 2) systematic component; and 3) link function. The first step is the matrix generation, then a linear regression will be applied to the matrix finding the best possible linear relation to fit the expected values. The systematic components are distributed on the line while random components are outside of the line. The link function will identify a relationship between the linear predictor and the distribution of the random components. It aims to explain why some values are following a linear repartition while others are not. This model performs well when using data that are already explained on unexplained data. The accuracy might, however, be hindered by outliers that may introduce misinterpretation. This model is used in the icluster + tools (Mo et al., 2013; Song et al., 2022).

Bayesian integrative clustering: the objective of this approach is to capture the major variations of multiple omics datasets, using a reduction of the high-dimensional space to a low-dimensional subspace. This could be vulgarized into a compaction of the data. The algorithm will extract the principal variations of the datasets to integrate matrices of different dimensions to a single matrix called Z with n x k dimensions, following the rule of multivariate normal distribution. A joint integrative clustering will capture relevant features to individualize distinct clusters across omics datasets.

TABLE 2 Publicly available cancer databases with their main characteristics.

Database	Multi-omics data available	Disease	References
TCGA (The Cancer Genome Atlas)	20,000 individual tumor samples	Cancers	(Weinstein et al., 2013; Tomczak et al., 2015)
	RNA-Seq, DNA-Seq, miRNA-Seq, SNV, CNV, DNA methylation, and RPPA, clinical data (treatment received and response to treatment), histological data		
TCIA (The Cancer Immunome Database)	8,000 tumor samples	Cancers	Charoentong et al. (2017)
	Genomic immune-related gene set, immune infiltrate, neoantigens, cancer antigens, HLA types, and tumor heterogeneity		
ICGC (International Cancer Genomics Consortium)	20,383 samples listed	Cancers	(Hudson et al., 2010; Australian Pancreatic Cancer Genome InitiativeBailey et al., 2016; Thompson et al., 2018)
	Somatic and germline whole genome sequencing, genomic variation data		
METABRIC (Molecular Taxonomy of Breast Cancer International Consortium)	2,503 breast tumor samples	Breast cancers	Curtis et al. (2012)
	Clinical traits, gene expression, SNP, and CNV.		
TARGET (Therapeutically Applicable Research to Generate Effective Treatments)	24 different types of pediatric cancers	Pediatric cancers	(Ma et al., 2018; Rajbhandari et al., 2018)
	Gene expression, miRNA expression, CNV, and DNA-seq data		
CRI (Cancer Research Institute) **iAtlas**	10,000 tumors samples	Cancers	(Thorsson et al., 2018; Eddy et al., 2020)
	Clinical data (immunotherapy responses and clinical phenotypes), genomics, immunomodulatory genes, neoantigens load		

DNA-seq: deoxyribo nucleic acid-sequencing.
RNA-seq: ribonucleic acid-sequencing.
miRNA-seq: micro ribonucleic acid-sequencing.
SNV: single nucleotide variant.
CNV: copy number variation.
SNP: single nucleotide polymorphisms.
DNA, methylation: deoxyribo nucleic acid methylation.
RPPA: reverse-phase proteomic arrays.
miRNA, expression: micro ribonucleic expression.
HLA: human leukocyte antigen.

The use of Baye's theorem enables the analysis of factors with various distributions and correlation among datasets. The probabilistic model is also permissive of missing values (Needham et al., 2007; Fang et al., 2018; Mo et al., 2018).

Bayesian factor analysis: this is a key component of the multi-omics factor analysis (MOFA) family tools used for multi-omics data integration. This unsupervised method infers principal component-based factors to decompose each matrix of M different omics components. The matrices will be transformed into a Z matrix of factors for each sample and M weight (W) matrices with features in rows and factors in columns, for each omics element. Downstream analysis will identify inference across Z and W matrices. Also, because the algorithm is based on a Bayesian framework, the distributions of the data are placed on the unobserved variables of the models and the algorithm will keep running until all the data have been characterized (Needham et al., 2007; Argelaguet et al., 2018; Min et al., 2018).

To finish, despite the numerous tools that are now available, the perfect tool does not exist. The continual development of computational methods necessitates systematic evaluation (benchmarking) of the omics data analyses tools and methods (Mangul et al., 2019). The major issue on this benchmarking is the lack of "gold standard" datasets, providing unbiased ground truth. This lack of gold standard hinders the possibility to establish generalizable benchmarks to test novel complex software (Mangul et al., 2019; Weber, 2019; Marx, 2020). Different notable benchmarks are available comparing: multi-omics and multi-view clustering algorithms (Rappoport and Shamir, 2018), multi-omics dimensionality reduction (Cantini et al., 2021), multi-omics for cancer subtyping (Duan et al., 2021),

and multi-omics survival prediction methods (Herrmann et al., 2021). Regardless of the use of the same dataset for comparison, the results obtained are not necessarily reproducible on another database and cannot be apply to the integration of omics that were not included in the initial dataset. The lack of a gold standard dataset impairs the success of the benchmarking (Krassowski et al., 2020).

Multi-omics database

The expansion and acceleration of NGS have generated a tremendous amount of data. In a collaborative effort, extensive omics databases have been created to stock and make those data publicly accessible to researchers, allowing for large meta-analysis for a tumor category or across cancer types. These datasets can host the sequencing output of different biological elements from bulk tumor or single-cell sequencing as well as clinical and treatment information. We will provide a summary of the main databases that are available for cancer research and detail the characteristics of each library with tumor types included and the biological and clinical contents provided (Table 2).

The Cancer Genome Atlas (TCGA) is the largest pan-cancer multi-omics database with clinical annotation allowing for large meta-analysis. TCGA is widely used by the research community, promoting new discoveries on tumor biology, evolution, and treatment specific biomarkers validated on meta-analysis across cancer types (Weinstein et al., 2013; Tomczak et al., 2015).

The Cancer Immunome Database (TCIA) is a new database of immunogenic analysis of NGS data from 20 different types of solid tumors derived from TCGA database. This database has served for the elaboration of a pan-cancer immunogenomic classification for checkpoint blockade sensitivity (Charoentong et al., 2017).

The International Cancer Genomics Consortium (ICGC) is the most ambitious biomedical efforts research since the human genome project. This database has permitted novel observation of cancer biology through the whole genome annotated alterations from 2,800 samples (Hudson et al., 2010; Australian Pancreatic Cancer Genome InitiativeBailey et al., 2016; Thompson et al., 2018).

The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) contains genomic data from breast cancers. This database helped improving the classification and subtyping of breast tumors and already led to the discovery of 10 subgroups (Curtis et al., 2012).

Therapeutically Applicable Research to Generate Effective Treatments (TARGET) is a clinically annotate database hosting genomics and transcriptomics data from pediatric tumors. This dataset empowered the description of the driving process of childhood cancers, and was used, for example, for characterizing the immune environment of pediatric cancers and its prognostic

impact (Ma et al., 2018; Rajbhandari et al., 2018; Sherif et al., 2021).

Cancer Research Institute (CRI) iAtlas is a database platform allowing the study of interaction between tumors and immune microenvironment, granting the possibility to explore immune response across genomics and clinical phenotypes. iAtlas (first version) allows researchers to explore these readouts and the relation between tumor types and immune response (Thorsson et al., 2018). CRI iAtlas now shares information about immunotherapy response useful for biomarker identification (Eddy et al., 2020).

The dramatic acceleration in data production and the policy for data sharing have increased the comprehension of cancer complexity and fostered the integration of interrelated biological elements of omics analysis in more and more complex computational models that we will show in detail in the following.

Multi-omics for greater accuracy

Multi-omics for a better classification of cancer types

Tumor heterogeneity remains a hurdle to understand tumor biology. For example, for a same tumor type, there is a wide variation in clinical evolution across population of patients, highlighting differences in tumor cells to progress or mutate (Lagoa et al., 2020; Wu et al., 2020). These differences are barriers to develop efficient therapies (Marusyk et al., 2012; Cyll et al., 2017; Marusyk et al., 2020). Cancer is relatively easy to classify as the classification is based on the histotype and site of origin of the tumor. However, the classification system has increased in complexity in the last decades with the introduction of genomics and molecular features to better account for the clinical evolution and foster the identification of histologic groups (Carbone, 2020). Therefore, omics data can be used to individualize tumor types. In this aim, a variety of bioinformatic tools have been developed to refine and accelerate sample classification.

One of these tools that allows the sample classification is a microarray logic analyzer (MALA). MALA is based on logic data mining algorithm. Its follows a three-step process: (Santucci et al., 2020) discrete cluster analysis, (Ferlay et al., 2020) selection of the most relevant (cluster of) genes, and (Dagogo-Jack and Shaw, 2018) logical formulas characterizing the samples (Weitschek et al., 2012). In the context of cancer research, a comparative study came out with the purpose to evaluate which tools could be the most accurate in treating multilevel omics data. MALA with sparse canonical correlation analysis (SCCA) and non-negative matrix factorization (NMF) were compared. When using the experimental data, MALA performed the best sample classification compared to the others. However, on a larger data set of simulated data, the efficiency of the model decreased (Pucher et al., 2019).

Multi-omics for cancer subtyping

To better understand cancer, researchers need to characterize the different cancer subtypes. Subtyping cancer is more complicated and remains one of the major challenges in cancer research. The identification of subtypes can provide an understanding of the underlying molecular mechanisms and thereby help design precise treatment strategies for efficient cancer management. Contrary to classification that is more histologic, the subtypes are influenced by oncogenic alterations and/or modifications in the gene/protein expression (Anderson et al., 2021). Molecular classification has partially elucidated tumor heterogeneity, however, different subtypes can be identified depending on different layers of biological elements: genomics, alterations, gene and protein expression profile as well as cellular composition (Skoulidis and Heymach, 2019). The main challenge is to be able to accurately analyze the large amount of data and to determine and isolate predictive patterns. Multi-omics provides a powerful approach to process, treat, and characterize the large quantity of data required for cancer subtyping (Menyhárt and Győrffy, 2021).

Cancer subtyping can be efficiently addressed by matrix factorization, such as joint and individual variation explained (JIVE) algorithm. This tool aims to individualize two types of structures: joint and individual structures. The former is the biological patterns of the samples that are shared between the different component types (*i.e.*, gene expression and miRNA), while the latter is intrinsic of one component and unrelated to others (Lock et al., 2013). One structure can interfere in finding a signal in the second structure and *vice versa*. JIVE was developed to distinguish these possible interfering effects by decomposing the datasets into a sum of three terms: (Santucci et al., 2020) low-rank approximation capturing joint variation across the biological components, (Ferlay et al., 2020) low-rank approximation capturing structured variation specific to a given component, and (Dagogo-Jack and Shaw, 2018) residual noise. It explores the information provided by a specific biological data type or by the interaction between several data for subtyping. A logical weakness of this method is its sensitivity to outliers. JIVE algorithm was tested on 234 glioblastoma samples, with an input of miRNA and gene expression matrices. It showed that the gene expression introduced more structured variation than miRNA and that joint structure variation between the gene expression and miRNA was more accurate to classify the biological subgroups. Overall, the multi-component integration improved the subclassification of glioblastoma samples.

In the optic of diseases subtyping, a growing diversity of tools based on matrix factorization were developed. As an example, non-negative factorization (NMF) consists in multiplication between the columns and the rows of a matrix. The input data sets are formed by a matrix called A, composed of N genes and M samples. Genes that regulate the expression of downstream genes will be identified and labeled k. In the second step, the matrix A will be split in two matrices: W composed of the N genes and k, and H composed of k and M samples. Finally, W and H matrices will be combined in a new matrix. The advantage of the NMF technics is to easily cluster gene and samples, but this is a time and memory consuming tool that does not handle negative input and is not designed to integrate multilayer components (Lee and Seung, 1999; Brunet et al., 2004; Zhang et al., 2012; Yang and Michailidis, 2015).

Briefly, NMF extensions were implemented to allow multi-omics profiling. Multi-omics integration with jNMF was compared to single omics to class clinical data from TCGA into different subgroups and outperformed the single element model. Moreover, the number of groups will depend on the clinical data used (Zhang et al., 2012). A second major update was intNMF, while jNMF only considers homogenous effects in WHI and intNMF considers heterogenous effects (Yang and Michailidis, 2015). As an example, breast cancer subtyping with multi-omics data integration of five biological components by intNMF refined the subclassification from four classical subsets to six unique clusters (Chalise and Fridley, 2017).

The last update of NMF is called network-based integrative (nNMF). nNMF involves a two-step process with network generation and integration of the network. To generate the network, a consensus matrix is built with a binary value attributed to each sample to reflect the connectivity between samples. Successive cycles are performed attributing a new binary value for every novel entry until the matrix is stabilized. The mean of the consensus matrix is made for each iterative cycle until the last cycle. This generates a consensus matrix for each data type integrated. The larger elements of the consensus matrix reflect the higher similarity between samples. After the networks are generated, samples can be considered as vertices and the consensus values as the edges. The network integration is based on the message passing theory (updating and combining network) that is processed on two ways: strong signals present in any data are conserved and consistent signals in multiple data are added up during an iterative process. Weak signals disappear while filtering out the noise. An advantage of nNMF is its ability to detect true cluster of small size with high reliability. Clinical application of nNMF demonstrated the capacity to establish novel clusters on head and neck squamous cell carcinoma, glioblastoma, and low-grade glioma data sets, suggesting a new comprehensive subtyping that eliminates previously unclassified samples (Chalise et al., 2020).

NMF and its extensions are based on matrix factorization algorithm that are efficient tools to treat a massive amount of data but are underperforming with missing values. To complement this part, iCluster family tools are good alternatives. iCluster is based on probabilistic matrices algorithm and iCluster is based on a Gaussian latent variable model. The basic concept of

iCluster is to jointly estimate the link between the data with a dimension reduction principle: data and features are clustered together to maximize the correlation between data types (Shen et al., 2009). iCluster was used to classify novel subtypes of esophageal cancers based on genomics, epigenomics, and transcriptomics data. The classification varied depending on the type of omics. Compared to previous classification, the samples were consistently classified into three groups with different biological traits and prognostic significance (Ma et al., 2021). It was also used in breast cancer to integrate DNA and RNA data leading to novel subtyping with noticeable clinical outcomes beyond classic expression profiling (Curtis et al., 2012).

The first extension of iCluster was iCluster+ (or iClusterplus), a model based on a generalized linear regression combined to the basic algorithm of iCluster. Compared to different models, iCluster + produces the best classification when integrating unknown datasets. It has been useful to construct two molecular subtypes and identified two core genes (*CNTN4* and *RFTN1*) for lung adenocarcinoma (Zhao et al., 2021). Two limitations to this tools can be pinpointed: it needs to test hundreds of values to tune the optimal solution parameters and there is no evaluation of statistical significance for a selected feature (Sathyanarayanan et al., 2020).

The most recent upgrade is iClusterBayes, which uses a Bayesian integrative clustering algorithm. The main advances of iClusterBayes are to overcome the limitations of iCluster + that were priorly exposed. This tool was tested on kidney cancer and glioblastoma and grants the possibility for tumor subtyping. However, iClusterBayes was not compared to another method, so that its capacity to discover novel subtypes could not be assessed (Mo et al., 2018).

SLIDE is a tool based on the concept of the multi-view of data, using a matrix factorization algorithm. SLIDE was developed in the continuity of JIVE, trying to investigate on shared and individual structure. Compared to JIVE, SLIDE allows the creation of partially shared scores in addition to the individual ones. SLIDE was able to upgrade breast cancer subtyping using gene, methylation, miRNA, and protein profiling (Gaynanova and Li, 2019).

The integration of multiple layers of interconnected biological elements through a wide palette of multi-omics tools available is a great opportunity to better classify biological-relevant cancer subtypes and to reduce unclassified samples. The choice of the algorithm should judiciously fit the characteristic of a given dataset to optimize the model performance.

Biomarker discovery

The tumor biology is intimately related to disease evolution and treatment response (Marusyk et al., 2020). The assimilation of clinical data as additional features to feed multilayer integrated models enables association between molecular subtypes and clinical outcome. Discovering biomarkers associated with prognosis and treatment sensitivity/resistance is a keystone for risk-group classification and therapeutic decision. The identification of treatment specific biomarkers is also granting the opportunities to provide therapies tailored to the biological trait of a specific tumor, opening the path for precision medicine. Machine learning and deep learning models, trained on Kaplan–Meier derived survival data, are powerful classifiers to predict the clinical evolution. In this part, we aim to illustrate how multi-omics integration alongside machine and deep learning approaches might facilitate biomarker discovery and guide treatment decision.

Application of multi-omics to biomarker discovery

Multi-omics tools have been developed to discover novel biomarkers in oncology. For example, jNMF allows biomarker discovery for prediction of drug response through pathway signature analysis. The identification of novel connections between tumor biology and drug response highlighted an association between BRAF inhibitors efficacy and *BRAF/MITF* overexpression in breast cancer (Fujita et al., 2018). Also, iClusterBayes demonstrated its ability to discover biomarkers, revealing the role of *MTAP/CDKN2A/2B* expression for PD-L1 blockade sensitivity with a proportional relation of Kaplan–Meier survival to the gene expression level (Mo et al., 2020).

Lemon-tree is another tool for multi-omics processing that runs a series of tasks that are self-contained step in the learning and clustering process. The workflow of the tools is as follow: ask biological question → preprocess data → clusterization → builds modules of cluster → compute score → results. Using this tool to discover biomarkers was operated on glioblastoma genes, looking at the amplification levels and copy-loss level of genes. The results show that genes that have copy number alteration (CNA) of glioblastoma oncogene *EGFR* and tumor suppressors *CDKN2A* and *PTEN*, but also novel candidates such as *KRIT1* and *PAOX* were assimilated with a worse prognosis (Bonnet et al., 2015). iProFun, is a method analyzing the "cascade effects" of the genes. It takes as input statistics associated with the data, aiming to detect the joint variation between each data. This study highlighted potential therapeutic candidates (*AKT1*, *KRT8*, and *MAP2*) in ovarian cancers. But it also demonstrated the role of *BIN2* in ovarian cancer, and how it can be a favorable survival outcome (Song et al., 2019). Finally, AMARETTO is used to identify driver genes by integrating genomics and epigenomics data. This is a three-step process, identifying candidate cancer driver genes, modeling effect on gene expression, and association of drivers with their

targets. In other words, it creates clusters depending on driver genes and expression of the genes. AMARETTO was able to highlight different driver genes such as *GPX2* for smoking induced cancer (lung squamous cell carcinoma), but also identified *OAS2* and *TRIM22* as modulator of the immune response (Champion et al., 2018).

Machine learning model for biomarker discovery

Machine learning are algorithms that can predict models using statistical methods based on training data. Algorithms can be trained in two ways: *via* supervised or unsupervised learning. Supervised learning relies on a labelled input dataset used to train and develop a function to predict the outcome on an experimental dataset. The algorithm attempts to find patterns relating the features to the given label. The second step is to integrate data test and to classify the validation data in the right label by following the identified patterns. The correlation between the given and predicted labels can be compared to assess the accuracy of the model. Unsupervised learning clusters the samples by identifying and regrouping different features following a similar pattern (Huang et al., 2015). The main difference of supervised learning is that the input data are not labelled. Machine learning approaches algorithms containing only one hidden layer.

A powerful tool developed for multi-omics integration is the artificial neural network (ANN). It can be used for machine learning and for deep learning. ANN is a simple approach to create an artificial model, composed of neurons organized in layers. This is composed of an input layer, hidden layer, and an output layer. The input layer corresponds to the dataset given to the algorithm and the output layer corresponds to the possible outputs of the algorithm, in this case, a classification of the data. The hidden layer is what defines the complexity of the algorithm, it represents possible pathways linking the input layer to the output layer. For the integration of multi-omics, different blocks, specific to each type of omics data, are created. Depending on the biological question, different machine learning algorithms can be preferred.

A vast diversity of algorithms are available for machine learning; random forest classifier and k-nearest neighbor are commonly supervised classifiers that are easily suitable for biomarker discovery. A random forest algorithm builds multiple decision trees that are randomly generated. Each decision tree provides a classification. The classification that is chosen in a majority of cases is used to classify each datapoint of the dataset (Tin Kam, 1995). K-nearest neighbor follows three main steps: 1) a reference dataset is clustered into the different classes that need to be distinguished, 2) the experimental data are integrated into the dataset, and 3) the experimental data are classified by calculating the distance to each defined class and

classifying each datapoint in the nearest cluster (Fixt and Hodges, 2022).

These different approaches have been successfully applied to biological data to improve prognoses prediction. For example, ANN has already been used to analyze the survival in two breast cancer datasets. The model was able to predict the prognoses (favorable or unfavorable) and relapse probability (Chi et al., 2022). In the second example, machine learning approaches were compared to histopathological grading to classify and predict patients' outcome in glioblastoma. In this test, a k-nearest neighbor-based ANN approach out scaled the histopathological grading for tumor classification and survival prediction. (Petalidis et al., 2008). In the third example, a random forest algorithm allowed the discovery of a novel prognostic biomarker in the Ewing Sarcoma. As an example, a lower Ki67 expression was associated to a better prognosis for the subset of samples with low-CD99 expression (Bühnemann et al., 2014). In the final example, a gene expression-based algorithm of k-nearest neighbor and random forest identified novel prognostic genes had increased the accuracy for the classification of the poorly defined group of soft tissue sarcomas. They were also able to annotate the samples in molecular subsets with potential therapeutic susceptibility (van IJzendoorn et al., 2019).

Deep learning for treatment guiding

Deep learning approaches are a subset of machine learning that contains multiple hidden layers organized in a network allowing for progressive learning (LeCun et al., 2015; Bi et al., 2019). Deep learning can be deep neural network (DNN) or convolutional neural network (CNN). DNN is similar to ANN with an increased number of hidden layers, and CNN is a class of ANN used for imaging analysis (Simonyan and Zisserman, 2015; Szegedy et al., 2015; He et al., 2016; Chollet, 2017).

Survival analysis learning with multi-omics neural network (SALMON) attempts to aggregate and simplify the gene expression data to enable prognosis prediction. Deep learning approaches typically introduce an extensive number of parameters from a limited sample set that can introduce data overfitting rendering the models ineffective. In comparison, SALMON favorizes the use of eigengene matrices of gene co-expression modules to feed the model. This model was compared to other survival prognostic models on breast cancer datasets, including DeepSurv and random survival forest. SALMON predictor reached a higher survival concordance index than other methods and showed that the integration of multi-omics data combined with clinical and demographical features can increase the prediction performance. This algorithm was used to develop an age-specific prognostic score for breast cancer (Huang et al., 2019).

CNN is very useful for deconvolutional observations of histological images used to classify and predict cancer prognoses. CNN was successful to detect and to predict

patient outcomes on haematoxylin-eosin-stained tumor tissue microarray with a higher accuracy than the visual prognostic prediction. For that, the image was cut into spots that were characterized with a pre-trained convolutional neural network (Bychkov et al., 2018). MesoNet is another tool developed to analyze larger images. To do so, the whole-slide images are subdivided in tiles that have a score assigned. Based on the score, the algorithm can select the most relevant tiles for the prediction. The cumulated score of the selected tiles informs the prediction of a patient's overall survival (Courtiol et al., 2019). These tools perform well to analyze cancer heterogeneity. To finish, survival convolutional neural network (SCNN) is a CNN-based algorithm combined to an integration tool. This model identifies visual patterns from regions of interest isolated in biopsies and relates them to patient's outcome to create a model. This comprehensive tool can also integrate genomics biomarkers. This technology was able to develop a better prognostic prediction tool compared to WHO genomics classification (Mobadersany et al., 2018).

DrugCell is a visible neural network for drug response prediction. The model is built with two branches. The first one modulates the hierarchical organization of molecular subsystems in human cells. Each subsystem, involving small protein complexes, is connected to larger pathways and to cellular functions assigned from a bank of artificial neurons. The input layer is the mutation status from genomics data and the output corresponds to the state of the whole cell based on the genotype. The second branch of the system is a CNN assessing the fingerprint of a drug. The output from the two branches of the model is combined in a single layer of neurons integrating the response of a genotype to a certain treatment. To test on the first side of the model, they used two large drug screening resources: the Cancer Therapeutics Response Portal (CTRP) v2 and the Genomics of Drug Sensitivity in Cancer (GDSC) database. Using these two resources, it covers 684 drugs and 1,235 cell lines. Each cell-line genotype was represented with a binary vector (mutated or non-mutated). The chemical structure of the drug was represented by an average of 81 activated bits in the Morgan fingerprint vector. Drug cell was trained to associate each genotype-drug paired with its corresponding drug–response curve. As a last point, drug cell was clinically tested stratify cancer patients depending on their response to treatment. They tested drug cell on clinical trial data from 221 estrogen receptor positive metastatic breast cancer patients and showed that the model was able to predict the response to mTOR and CDK4/6 inhibitors. Drug cell also includes a feature for predicting the sensitivity to drug combination (Kuenzi et al., 2020).

In a number of studies, DNN was used to predict prognoses. As an example, DeepSurv is a tool defined as a prognostic model. With the integration of time-dependent and treatment-sensitive survival data to the model, it was trained to provide treatment recommendations. DeepSurv was tested on four real-life datasets: Worcester Heart Attack Study (WHAS), the Study to Understand Prognoses Preferences Outcomes and Risks of Treatment (SUPPORT), the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC), and Rotterdam & the German Breast Cancer Study Group (GBSG). DeepSurv takes input as the patient's baseline data. The output is composed of one node estimating the log-risk function in the Cox model (estimation of the effect parameters without any consideration of the hazard function). As the second part, it uses a treatment recommender system. This system takes into account a group of patients assigned to a specific treatment group to predict the log-risk of a given treatment. Using the same base line hazard functions, it is possible to calculate the personal risk-ratio of prescribing one treatment and then, determine a treatment recommendation algorithm. To assess the possibility of a treatment recommendation system, DeepSurv was trained on the Rotterdam tumor bank to build the recommendation system that was then tested on GBSG. The treatment recommendation system demonstrated that, on a real clinical dataset, the patient's survival would be increased by following the treatment recommendation of the algorithm. In comparison, a random survival forest-based recommendation system was not able to improve the patient's outcome (Katzman et al., 2018). To conclude, DeepSurv provides strong modeling capabilities with the ability of guiding the therapeutic decision and bringing the clinician closer to computer-assisted technology for precision medicine in oncology.

Other examples of machine and deep learning models applicable for biomarker discovery are detailed in a recent review articles devoted to this subject (Nicora et al., 2020) (Zhu et al., 2020). So far, these powerful algorithms have been used to successfully develop models for prognosis and treatment response prediction. However, these cutting-edge technologies are not yet ready for personalized treatment recommendation in the perspective of prospective clinical implementation.

New hope in omics: deciphering host–tumor interplay

Novel fields of interest have emerged to unravel host–tumor interactions (Hui, 1989) through the immunologic cancer response (Nisar et al., 2020) (Binnewies et al., 2018) and to understand how tumor cells can trick the host to hijack its defensive system and favor cancer immune evasion. The comprehension of the immune escape mechanisms opened the path to immune-mediated therapeutic opportunities. ICB are molecules that restore the immune system's ability to recognize and eliminate tumor cells such as PD1/PDL1 (program cell-death protein one and its ligand) or CTLA4 (cytotoxic T-lymphocyte associated protein 4) inhibitors have revolutionized the treatment of some adult cancers (Nishino et al., 2017). However, not all the patients with a same disease will respond to these treatments and anticipating the patients who

will be responders remains challenging. Indeed, ICB efficacy is highly dependent of the constitution of the tumor cell immunogenicity, the host immune recognition and, ultimately, the TiME (Bonaventura et al., 2019). The advances in NGS and the growing number of high-throughput tumor sequencing datasets alongside the constant improvement of computational analyzes opened the way to new disciplines like immunomics and microbiomics.

ICB biomarker identification

The host-tumor interaction comprises a succession of interrelated steps facilitating or refraining an efficient immune recognition of tumor cells. These steps are dependent of the tumor biology, the host ability to recognize the tumor, and the mechanisms that will promote immune inflammation or immune evasion. Briefly, the somatic mutations of tumors result in the expression of neoantigens that can be exposed to and recognized by the host antigen presenting cells (APCs). This will activate an immune cascade responsible for the activation and recruitment of effector T cells in the tumor (Chen and Mellman, 2017). These primary steps determine the infiltration of the tumor by effector immune cells: the tumor infiltrated lymphocytes (TILs). The anti-tumor activity of TILs is influenced by the presence of stimulatory or inhibitory signaling. PDL1, expressed by tumor and immune cells, binds to the PD1 receptor on T cell and generates an inhibitory signal leading to T-cell exhaustion and loss of activity (Chen and Mellman, 2017). To be fully efficient, ICBs needs the tumor to be infiltrated by inflamed T cell rending possible to develop biomarkers of efficacy (Bonaventura et al., 2019).

The identification of biomarkers was first made possible by immunohistochemistry (IHC) assay that has the advantage of spatial annotation but is limited by the sparse number of features that can be assessed. A classification in three IHC immune states for tumor infiltrated lymphocytes (TILs) density is usually admitted with immune inflamed (or "hot"), immune excluded, and immune deserted (or "cold") tumors (PMID: 30,179,157). Few biomarkers are so far broadly recognized. Some from the IHC like tumor infiltrated lymphocytes (TILs) density, mostly for CD8, and the protein expression of PD-L1 in the tumor. Other biomarkers derive from NGS test: tumor mutation burden (TMB) and the level of expression of interferon-gamma (*IFNγ*) or *INFγ* gene expression profile signature (Ayers et al., 2017; Nishino et al., 2017). These biomarkers are known to be independent predictors of response to ICB, however the predictive value of cumulative biomarkers remains partially solved, so that the integration of multilayers omics analysis for dissecting the immune environment through immunomics was brought to the fore.

Immunomics for TiME profiling

Immunomics is the part of omics integration that aims to comprehend the host-tumor interaction and to profile the TiME. Integrative immunomics allows for quantitative, functional, spatial, and clonal annotation of the immune environment to comprehend the host/cancer interaction and discover biomarkers for ICB efficacy.

To integrate and study immune cell-host interactions, novel bioinformatic tools were developed. The quantification and enumeration of immune cell infiltrate can be deduced from bulk-tumor RNAseq data by deconvolution-based tools, such as MCP-counter (Becht et al., 2016) and CIBERSORT (Newman et al., 2019), or through gene expression scores (Danaher et al., 2017). It is also possible to use computational analyzes to infer the somatic mutations and the neoantigens presented by the tumor cells as well as the immunogenicity of these neoantigens, for example, the TMB assesses the mutational load of a tumor (Endris et al., 2019). The enumeration of intra-tumor T-cell and B-cell clonotype expansion, through T-cell receptor (TCR) and B-cell receptor (BCR) rearrangement, are particularly interesting as a surrogate of the immune reactivity potentially induced by the tumor. Different tools exist to extract the T-cell and B-cell clonotype repertoire from genomics or transcriptomics data, such as MiXCR (Bolotin et al., 2015) and immuneDB (Rosenfeld et al., 2018). These methods align VDJ sequences, quantify them, and sort them in distinct clonotypes. With the development of these tools, high-throughput NGS dataset can be analyzed through the spectre of immunomics and for discovering new biomarkers of ICB efficacy. We will, herein, present some of the latest studies of immunomics integration.

A pan-cancer analysis, from the publicly available TARGET dataset, aimed to characterize and classify immune subsets of pediatric solid cancers based on transcriptomics and survival data (Sherif et al., 2021). The gene expression profile and gene set enrichment analysis (ssGSEA) individualized six distinct immune groups with diagnostic and prognostic association. The study used the data from 408 samples from five pediatric cancers: neuroblastoma, osteosarcoma, and three kinds of renal cancers.

The immune infiltrate was first assessed by the immunologic constant of rejection (ICR) method (Thorsson et al., 2018) in three classes from low- to high-immune score. Kidney rhabdoid tumors had the higher immune score and Wilms tumor the lowest. Correlation clustering of ssGSEA-based immune signatures identified five main modules: interferon-gamma (IFN-G) and tumor growth factor beta (TGF-B) signaling, macrophages, lymphocytes, and wound healing. These five modules were used for the identification of six immune subtypes: T-cell helper 2 (Th2) dominant, inflammatory, immunologically quiet, wound healing dominant, macrophages dominant, and lymphocyte suppressed subtypes. Each immune subtype was composed by a specific immune

infiltrate by CIBERSORT enumeration, that mirrored the immune signature.

Tumor types were unevenly distributed between the immune groups, and thus highlighting the association of the immune phenotype and the diagnosis. Furthermore, the immune subtyping was correlated with the prognosis, inflammatory subtype having the best clinical outcome while wound healing dominant subtype had the worse survival. The clinical impact was observed across cancers and within a same cancer type. This work demonstrates that access to NGS, well clinically annotated large tumor database, and immunogenomics integration help understand the interplay of host immune system and tumor cells. The immune landscape characterization can also improve tumor classification and prognostication as shown in this pediatric database.

A systematic meta-analysis of tumor and microenvironmental biomarkers for ICB sensitivity has investigated the relative impact of independent and combined biomarkers. An immunomics analysis from bulked-tumor transcriptomics and genomics was performed on a dataset from seven cancer types from 1,008 adult patients (CPI1000+) treated with immune checkpoint blockade (ICB) (Litchfield et al., 2021). The goal of this work was to select a large panel of biomarkers, aggregated by a systematic literature search, to assign a Z score for each of them and test their prediction impact for ICB efficacy in a large pan-cancer population. The different biomarkers explored the T-cell response, the mechanisms for immune evasion and infiltration, and the host factors.

In univariate analysis, clonal and total TMB and *CXCL9* expression, a CD8 attractive chemokine, were the strongest predictors for ICB response, followed by *CD8A* expression, T-cell inflamed *INFG* gene expression, and *CD274* expression. Due to the high prevalence of TMB in the literature, researchers decided to subdivide the somatic mutation by mutations fitting in an immunogenic signature. Some signatures like dinucleotide variants, that are source of amino-acid changes and generate immunogenic epitopes, ultraviolet (UV), or tobacco mutation signatures also came out as significant determinant of ICB response. In somatic copy number analysis, two copy number anomalies were identified as positive or negative predictor of response. Surprisingly, host factors like the loss of HLA heterogeneity or HLA subtypes did not show any impact on ICB sensitivity. They also performed single-cell RNA sequencing of a reactive CD8 TILs from patients to identify T-cell intrinsic markers of ICB sensitivity and highlighted *CXCL13* and *CCR5* gene involvement. The immune cell enumeration or BCR/TCR clonality assay were not assessed in this study.

A machine learning model for multivariate analysis including the significant determinant of response was a better predictor of ICB response than TMB alone in the initial cohort and three external independent validation datasets. Also, a two-parameter

biomarker model combining clonal TMB and *CXCL9* expression had a better prediction accuracy than clonal TMB alone, yet inferior to the multivariate model. The integration of multiparameter biomarkers could explain approximatively 60% of the response to ICB in the different cancer types.

This study showed that multilayers integration of immunogenomics data and multiparameter models can improve the prediction of biomarker to anticipate the response to immunotherapy. It is open the opportunity to mechanistically solve host-tumor interaction and to understand how ICB remodel the tumor environment.

Others studies of multilayer immunomics analysis depicted the role of B cell in promoting ICB efficacy (Anagnostou et al., 2020; Helmink et al., 2020). In Helmink's study, conducted on multiple cohorts of melanoma and renal cell carcinoma treated with ICB, for comparison between responders and non-responders, or immunotherapy naïve to test the prognostic impact of B-cell infiltrate. The input data comprised RNAseq, deconvolution immune cell enumeration, BCR clonality, single-cell RNA sequencing, and mass cytometry (CyTOF) for functional analysis and histological evaluation for spatial annotation. The gene expression profile showed an increase in the activation marker genes of B cells (*IFNG*, *MZB1*, *JCHAIN,* and *IGLL5*) in ICB responders. Using TRUST-algorithm for BCR clonotype enumeration, they observed an increase of clonal counts for heavy and light chains and in BCR diversity in responders. MCP-counter deconvolution algorithm also confirmed the enrichment of B cells in responders compared to non-responders. To confirm the role of B cell in ICB activity, they used IHC to assess B-cell density and to interrogate the spatial repartition of B cell and they performed functional study. The IHC confirmed an increase in B-cell density in responders and revealed a spatial organization in tertiary lymphoid structure (TLS) where CD20[+] B cells colocalize with CD4[+], CD8[+], and FOXP3+ T cells. A previous study similarly showed a correlation between B-cell signature and increased the expression of *CD8A* and CD8[+] T-cell infiltration (Griss et al., 2019). Functional characterization of tumor B cell by single-cell RNA-seq and CyTOF, first confirmed the B-cell enrichment, but also decipher a unique immune activity of B cell (increased activity in CXCR4 signaling, cytokine receptor interaction, and chemokine signaling pathways) and a switch in immune activated CXCR3+ memory B cells in responders. The use of single-cell RNAseq and spatial omics in cancer analyze allowed, in this case, to discover unsuspected novel interaction. This comprehensive approach has revealed the structural role of B cell and tertiary lymphoid structures in responses to ICB treatment.

Multi-omics integration for immune characterization has dramatically furthered the identification of biomarkers raising the possibility of personalizing ICB treatment based on the tumor immune constitution.

Microbiomics as a modulator of the immune environment

The microbiome is the community of microbial species that inhabit a human body (Bhatt et al., 2017). The study of interactions between cancer and the microbiome has gained popularity in the past few years, showing the interplay between cancer features and species colonizing the intestinal flora. The intestinal flora can, indeed, play a role on cancer (Sepich-Poore et al., 2021) by influencing the risk of cancer apparition (Song et al., 2020b) and cancer immune response, it also may be useful for cancer detection (Chen et al., 2021). Theses interactions can lead to mucosal inflammation or systemic metabolic/immune dysregulation and modulate immune responses by altering anti-cancer immunity and response to therapy (Bhatt et al., 2017; Gopalakrishnan et al., 2018; Riquelme et al., 2019; Dohlman et al., 2021; Jackson et al., 2022). In this part, we will show that immunomics and microbiomics data should be integrated into classic omics analysis to master cancer complexity.

There are different sequencing techniques to elucidate the microbiome composition. The first one is the sequencing of the bacterial 16 S ribosomal RNA (rRNA) that is abundant and specific to each species. The 16 S rRNA genes are isolated and amplified from microbiome containing samples and then sequenced for species characterization. The downside of this technique is the need of high accuracy in the primer for amplification (Janda and Abbott, 2007; Pei et al., 2010). A second technique, majorly used in microbiomics, is shotgun sequencing that uses the taxonomic, functional, and genomics profile of the bacteria to deduct the microbiome composition (Quince et al., 2017). The study of microbiome can also be done using classic omics techniques like transcriptomics (metatranscriptomics), proteomics (metaproteomics), and metabolomics used to detect dysregulation of genes, proteins, and metabolites (Franzosa et al., 2014; Bikel et al., 2015; Singhal et al., 2015; Daliri et al., 2017).

Lately, in the aim of cancer treatment, the study of microbiome allowed significant discoveries. It was first stated that antibiotics can disrupt the activity of immunotherapy inducing loss of response to immune checkpoint blockade (ICB). In murine model of melanoma, it was shown that the use of different antibiotics decreases the response of PD-1 treatment compromising its tumor effect (Routy et al., 2018). The same effect was observed in patients with non-small cell lung cancer (NSCLC), renal cell carcinoma (RCC), and urothelial carcinoma. The use of those antibiotics negatively impacted their overall survival and progression-free survival (Derosa et al., 2022) (Routy et al., 2018). This observation suggested that a balanced microbiota is necessary for inducing and maintaining the tumor immune response and prompted further evaluation of microbiome in patients treated with ICB. Microbiome exploration by metagenomics in NSCLC patients treated with ICB confirmed the impact of intestinal flora,

showing an enrichment of some species in responders, notably Akkermansia muciniphila (AkM) and Firmicutes, and in non-responder (Prevotella, *Clostridium* species) (Derosa et al., 2018; Routy et al., 2018). A recent work confirmed the role of AkM in response to ICB and introduced the notion relative abundance of bacteria in cancer (Derosa et al., 2022). They demonstrated that a high abundance of AkM can improve the patient outcomes in NSCLC, but if present in high quantity, AkM will be deleterious for ICB response. This raised the hypothesis that restoring microbiota equipoise could restore the response. This was assessed by fecal microbiota transplantation of responder flora in non-responders that was able to shift from unsensitive to sensitive phenotype and restore ICB efficacy. After operating the transplantation, 16 S RNA sequencing revealed that a higher abundance of Veillonellaceae family and poor in *Bifidobacterium bifidum* increased the sensitivity to the treatment (Baruch et al., 2021).

The role of microbiomics for modeling the host-tumor interaction and the immune response to cancer is now established, however, microbiomics has not yet been routinely implemented in multi-omics analysis but is probably a key element of cancer biology. Considering the high plasticity of the microbiome constitution, longitudinal analysis of the flora should be preferred (Turnbaugh et al., 2009; Khoruts et al., 2010; Spencer et al., 2011; Kong et al., 2012). Most importantly, the gut microbiota is highly intricated with the host immune reaction to the tumor warranting its integration in the models of tumor immune environment analysis.

Improving multi-omics integration in the context of immunomics

As we have shown, combining multiple biological elements of the tumor and of the host environment is an extremely powerful way to anatomize the TiME and to discover new biomarkers. However, there is no standardized method for immunomics and most of the computational integration tools are home made. To improve the generalization of the observation and to accelerate the possibility of patient selection for personalized therapy, standardization of the method is urging.

In that sense, an available R package for computational tool for immuno-oncology biological research (IOBR) has been developed to comprehensively combine and interpret multi-omics data in the context of immuno-oncology (Zeng et al., 2021). IOBR has been built to easily integrate whole exome and RNA sequencing from bulk tumor as well as single-cell RNA sequencing and long non-coding RNA data. It consists of a four-module pipeline with a signature/deconvolution, phenotype, mutation, and model construction. The signature module can identify immune or tumor specific signatures through expression gene profile and deconvolution estimate of the immune infiltrate. The deconvolution part directly integrates CIBERSORT (Newman et al., 2019), TIMER (Li et al., 2020), MCP-Counter (Becht et al.,

2016), xCELL (Aran et al., 2017), EPIC (Racle and Gfeller, 2020), and quanTIseq (Finotello et al., 2019). The phenotype module tests a large set of immune and non-immune-related published phenotypes, and the mutation part is able to determine association between different gene alterations and signatures. Finally, this tool has a model construction module that provides robust biomarker identification and model construction from the prior modules. The utilization of standardized ready-to-use package for onco-immunomics analysis is crucial to hasten biomarker discovery and test their inter-cohort reproducibility.

Future directions to fully characterize the immune environment and its dynamic would be to implement microbiomics data to such model and to construct longitudinal models with the introduction of time dependent models (Bodein et al., 2022). The incorporation of spatial annotation and single-cell omics to explore the cancer architecture at the cell level in all its heterogeneity and to complement with functional analysis offers a new venue to elucidate the TiME (Roh et al., 2018; Helmink et al., 2020; Zheng et al., 2021).

Conclusion

Cancer is a complex and highly heterogenous disease that is yet partially solved. Despite the molecular characterization made possible by the advances of NGS technologies, many of the underlying oncogenic mechanisms remain puzzling. The integration of the multilayers of the omics elements is an avenue to further elucidate cancer biology. In this review, we highlighted the challenges of computational modelization of large multi-omics datasets and we provided some clues for how to overcome these barriers. We presented some innovative bioinformatic tools developed to enlighten the implication of all the interconnected biological elements of cancer and showed how determinant they are to refine the disease subtyping and classification, and to discover biomarkers. Multi-omics integration has also heightened the field of immunomics and microbiomics, and thus has dramatically accelerated the identification of robust biomarkers for ICB efficacy toward the development of tailored immunotherapy.

The constant modernization of the models endows analyses of increasingly larger datasets with a growing number of components. Collaborative effort for high-quality and clinically well-annotated databases, combining all the elements of high-throughput sequencing, is crucial to feed the models. Finally, the standardization of the methods would aid in replicating and confirming the results to increase the global knowledge and, ultimately, improve cancer treatments.

Author contributions

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alyass, A., Turcotte, M., and Meyre, D. (2015). From big data analysis to personalized medicine for all: Challenges and opportunities. *BMC Med. Genomics* 8 (1), 33. doi:10.1186/s12920-015-0108-y

Anagnostou, V., Bruhm, D. C., Niknafs, N., White, J. R., Shao, X. M., Sidhom, J. W., et al. (2020). Integrative tumor and immune cell multi-omic analyses predict response to immune checkpoint blockade in melanoma. *Cell Rep. Med.* 1 (8), 100139. doi:10.1016/j.xcrm.2020.100139

Anderson, P., Gadgil, R., Johnson, W. A., Schwab, E., and Davidson, J. M. (2021). Reducing variability of breast cancer subtype predictors by grounding deep learning models in prior knowledge. *Comput. Biol. Med.* 138, 104850. doi:10.1016/j.compbiomed.2021.104850

Aran, D., Hu, Z., and Butte, A. J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 18 (1), 220. doi:10.1186/s13059-017-1349-1

Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J. C., et al. (2018). Multi-omics factor Analysis—A framework for unsupervised integration of multi-omics data sets. *Mol. Syst. Biol. [Internet]* 14 (6), 1. doi:10.15252/msb.20178124

Arnaout, R. A., Prak, E. T. L., Schwab, N., and Rubelt, F. (2021). The future of blood testing is the immunome. *Front. Immunol.* 12, 626793. doi:10.3389/fimmu.2021.626793

Australian Pancreatic Cancer Genome InitiativeBailey, P., Chang, D. K., Nones, K., Johns, A. L., Patch, A. M., et al. (2016). Genomic analyses identify

molecular subtypes of pancreatic cancer. *Nature* 531 (7592), 47–52. doi:10.1038/nature16965

Ayers, M., Lunceford, J., Nebozhyn, M., Murphy, E., Loboda, A., Kaufman, D. R., et al. (2017). IFN-γ–related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Invest.* 127 (8), 2930–2940. doi:10.1172/JCI91190

Barriga, V., Kuol, N., Nurgali, K., and Apostolopoulos, V. (2019). The complex interaction between the tumor micro-environment and immune checkpoints in breast cancer. *Cancers* 11 (8), 1205. doi:10.3390/cancers11081205

Baruch, E. N., Youngster, I., Ben-Betzalel, G., Ortenberg, R., Lahat, A., Katz, L., et al. (2021). Fecal microbiota transplant promotes response in immunotherapy-refractory melanoma patients. *Science* 371 (6529), 602–609. doi:10.1126/science.abb5920

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 17 (1), 218. doi:10.1186/s13059-016-1070-5

Berlanga, P., Pierron, G., Lacroix, L., Chicard, M., Adam de Beaumais, T., Marchais, A., et al. (2022). The European MAPPYACTS trial: Precision medicine program in pediatric and adolescent patients with recurrent malignancies. *Cancer Discov.* 12 (5), 1266–1281. doi:10.1158/2159-8290.CD-21-1136

Bertolazzi, P., Felici, G., Festa, P., and Lancia, G. (2008). Logic classification and feature selection for biomedical data. *Comput. Math. Appl.* 55 (5), 889–899. doi:10.1016/j.camwa.2006.12.093

Bhatt, A. P., Redinbo, M. R., and Bultman, S. J. (2017). The role of the microbiome in cancer development and therapy. *Ca. Cancer J. Clin.* 67 (4), 326–344. doi:10.3322/caac.21398

Bi, Q., Goodman, K. E., Kaminsky, J., and Lessler, J. (2019). What is machine learning? A primer for the epidemiologist. *Am. J. Epidemiol.* 188, 2222–2239. doi:10.1093/aje/kwz189

Bikel, S., Valdez-Lara, A., Cornejo-Granados, F., Rico, K., Canizales-Quinteros, S., Soberón, X., et al. (2015). Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: Towards a systems-level understanding of human microbiome. *Comput. Struct. Biotechnol. J.* 13, 390–401. doi:10.1016/j.csbj.2015.06.001

Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F., Merad, M., et al. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* 24 (5), 541–550. doi:10.1038/s41591-018-0014-x

Bodein, A., Scott-Boyer, M. P., Perin, O., Lê Cao, K. A., and Droit, A. (2022). timeOmics: an R package for longitudinal multi-omics data integration. *Bioinformatics* 38 (2), 577–579. doi:10.1093/bioinformatics/btab664

Bolotin, D. A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I. Z., Putintseva, E. V., et al. (2015). MiXCR: Software for comprehensive adaptive immunity profiling. *Nat. Methods* 12 (5), 380–381. doi:10.1038/nmeth.3364

Bonaventura, P., Shekarian, T., Alcazer, V., Valladeau-Guilemond, J., Valsesia-Wittmann, S., Amigorena, S., et al. (2019). Cold tumors: A therapeutic challenge for immunotherapy. *Front. Immunol.* 10, 168. doi:10.3389/fimmu.2019.00168

Bonnet, E., Calzone, L., and Michoel, T. (2015). Integrative multi-omics module network inference with Lemon-Tree. *PLoS Comput. Biol.* 11 (2), e1003983. doi:10.1371/journal.pcbi.1003983

Brunet, J. P., Tamayo, P., Golub, T. R., and Mesirov, J. P. (2004). Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci. U. S. A.* 101 (12), 4164–4169. doi:10.1073/pnas.0308531101

Bühnemann, C., Li, S., Yu, H., Branford White, H., Schäfer, K. L., Llombart-Bosch, A., et al. (2014). Quantification of the heterogeneity of prognostic cellular biomarkers in ewing sarcoma using automated image and random survival forest analysis. *PLoS ONE* 9 (9), e107105. doi:10.1371/journal.pone.0107105

Bychkov, D., Linder, N., Turkki, R., Nordling, S., Kovanen, P. E., Verrill, C., et al. (2018). Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci. Rep.* 8 (1), 3395. doi:10.1038/s41598-018-21758-3

Campbell, J. D., Yau, C., Bowlby, R., Liu, Y., Brennan, K., Fan, H., et al. (2018). Genomic, pathway network, and immunologic features distinguishing squamous carcinomas. *Cell Rep.* 23 (1), 194–212. doi:10.1016/j.celrep.2018.03.063

Cantini, L., Zakeri, P., Hernandez, C., Naldi, A., Thieffry, D., Remy, E., et al. (2021). Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer. *Nat. Commun.* 12 (1), 124. doi:10.1038/s41467-020-20430-7

Canzler, S., Schor, J., Busch, W., Schubert, K., Rolle-Kampczyk, U. E., Seitz, H., et al. (2020). Prospects and challenges of multi-omics data integration in toxicology. *Arch. Toxicol.* 94 (2), 371–388. doi:10.1007/s00204-020-02656-y

Carbone, A. (2020). Cancer classification at the crossroads. *Cancers* 12 (4), 980. doi:10.3390/cancers12040980

Chalise, P., and Fridley, B. L. (2017). Integrative clustering of multi-level 'omic data based on non-negative matrix factorization algorithm. *PLoS One* 12 (5), e0176278. doi:10.1371/journal.pone.0176278

Chalise, P., Ni, Y., and Fridley, B. L. (2020). Network-based integrative clustering of multiple types of genomic data using non-negative matrix factorization. *Comput. Biol. Med.* 118, 103625. doi:10.1016/j.compbiomed.2020.103625

Champion, M., Brennan, K., Croonenborghs, T., Gentles, A. J., Pochet, N., and Gevaert, O. (2018). Module analysis captures pancancer genetically and epigenetically deregulated cancer driver genes for smoking and antiviral response. *EBioMedicine* 27, 156–166. doi:10.1016/j.ebiom.2017.11.028

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* 18 (1), 248–262. doi:10.1016/j.celrep.2016.12.019

Chauvel, C., Novoloaca, A., Veyre, P., Reynier, F., and Becker, J. (2020). Evaluation of integrative clustering methods for the analysis of multi-omics data. *Brief. Bioinform.* 21 (2), 541–552. doi:10.1093/bib/bbz015

Chen, D. S., and Mellman, I. (2017). Elements of cancer immunity and the cancer-immune set point. *Nature* 541 (7637), 321–330. doi:10.1038/nature21349

Chen, H., Ma, Y., Liu, Z., Li, J., Li, X., Yang, F., et al. (2021). Circulating microbiome DNA: An emerging paradigm for cancer liquid biopsy. *Cancer Lett.* 521, 82–87. doi:10.1016/j.canlet.2021.08.036

Chi, C. L., Street, W. N., and Wolberg, W. H. (2022). *Application of artificial neural network-based survival analysis on two breast cancer datasets*, 5.

Chollet, F. (2017). "Xception: Deep learning with depthwise separable convolutions," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet], Honolulu, HI, cité 10 janv 2022 (IEEE), 1800–1807. Disponible sur: http://ieeexplore.ieee.org/document/8099678/.

Chu, J., Sun, N., Hu, W., Chen, X., Yi, N., and Shen, Y. (2022). The application of bayesian methods in cancer prognosis and prediction. *Cancer Genomics Proteomics* 19 (1), 1–11. doi:10.21873/cgp.20298

Courtiol, P., Maussion, C., Moarii, M., Pronier, E., Pilcer, S., Sefta, M., et al. (2019). Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nat. Med.* 25 (10), 1519–1525. doi:10.1038/s41591-019-0583-3

Curtis, C., Shah, S. P., Chin, S. F., Turashvili, G., Rueda, O. M., Dunning, M. J., et al. (2012). The genomic and transcriptomic architecture of 2, 000 breast tumours reveals novel subgroups. *Nature* 486 (7403), 346–352. doi:10.1038/nature10983

Cyll, K., Ersvær, E., Vlatkovic, L., Pradhan, M., Kildal, W., Avranden Kjær, M., et al. (2017). Tumour heterogeneity poses a significant challenge to cancer biomarker research. *Br. J. Cancer* 117 (3), 367–375. doi:10.1038/bjc.2017.171

Dagogo-Jack, I., and Shaw, A. T. (2018). Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* 15 (2), 81–94. doi:10.1038/nrclinonc.2017.166

Daliri, E. B. M., Wei, S., Oh, D. H., and Lee, B. H. (2017). The human microbiome and metabolomics: Current concepts and applications. *Crit. Rev. Food Sci. Nutr.* 57 (16), 3565–3576. doi:10.1080/10408398.2016.1220913

Danaher, P., Warren, S., Dennis, L., D'Amico, L., White, A., Disis, M. L., et al. (2017). Gene expression markers of tumor infiltrating leukocytes. *J. Immunother. Cancer* 5 (1), 18. doi:10.1186/s40425-017-0215-8

de Anda-Jáuregui, G., and Hernández-Lemus, E. (2020). Computational oncology in the multi-omics era: State of the art. *Front. Oncol.* 10, 423. doi:10.3389/fonc.2020.00423

Derosa, L., Hellmann, M. D., Spaziano, M., Halpenny, D., Fidelle, M., Rizvi, H., et al. (2018). Negative association of antibiotics on clinical activity of immune checkpoint inhibitors in patients with advanced renal cell and non-small-cell lung cancer. *Ann. Oncol.* 29 (6), 1437–1444. doi:10.1093/annonc/mdy103

Derosa, L., Routy, B., Thomas, A. M., Iebba, V., Zalcman, G., Friard, S., et al. (2022). Intestinal Akkermansia muciniphila predicts clinical response to PD-1 blockade in patients with advanced non-small-cell lung cancer. *Nat. Med.* 28 (2), 315–324. doi:10.1038/s41591-021-01655-5

Devarajan, K. (2008). Nonnegative matrix factorization: An analytical and interpretive tool in computational biology. *PLoS Comput. Biol.* 4 (7), e1000029. doi:10.1371/journal.pcbi.1000029

Dohlman, A. B., Arguijo Mendoza, D., Ding, S., Gao, M., Dressman, H., Iliev, I. D., et al. (2021). The cancer microbiome atlas: A pan-cancer comparative analysis to distinguish tissue-resident microbiota from contaminants. *Cell Host Microbe* 29 (2), 281–298. e5. doi:10.1016/j.chom.2020.12.001

Duan, R., Gao, L., Gao, Y., Hu, Y., Xu, H., Huang, M., et al. (2021). Evaluation and comparison of multi-omics data integration methods for cancer subtyping. *PLoS Comput. Biol.* 17 (8), e1009224. doi:10.1371/journal.pcbi.1009224

Eddy, J. A., Thorsson, V., Lamb, A. E., Gibbs, D. L., Heimann, C., Yu, J. X., et al. (2020). CRI iAtlas: An interactive portal for immuno-oncology research. *F1000Res* 9, 1028. doi:10.12688/f1000research.25141.1

Endris, V., Buchhalter, I., Allgäuer, M., Rempel, E., Lier, A., Volckmar, A., et al. (2019). Measurement of tumor mutational burden (TMB) in routine molecular diagnostics: *In silico* and real-life analysis of three larger gene panels. *Int. J. Cancer* 1, 2303–2312. doi:10.1002/ijc.32002

Fang, Z., Ma, T., Tang, G., Zhu, L., Yan, Q., Wang, T., et al. (2018). Bayesian integrative model for multi-omics data with missingness. *Bioinformatics* 34 (22), 3801–3808. doi:10.1093/bioinformatics/bty775

Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., et al. (2020). *Cancer statistics for the year 2020: An overview*, 12.

Finotello, F., Mayer, C., Plattner, C., Laschober, G., Rieder, D., Hackl, H., et al. (2019). Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med.* 11 (1), 34. doi:10.1186/s13073-019-0638-6

Fixt, E., and Hodges, J. L. (2022). Discriminatory analysis. *Nonparametric Discrim. Consistency Prop.* 11, 1.

Franzosa, E. A., Morgan, X. C., Segata, N., Waldron, L., Reyes, J., Earl, A. M., et al. (2014). Relating the metatranscriptome and metagenome of the human gut. *Proc. Natl. Acad. Sci. U. S. A.* 111 (22), E2329–E2338. doi:10.1073/pnas.1319284111

Fujita, N., Mizurai, S., Murakami, K., and Nakai, K. (2018). Biomarker discovery by integrated joint non-negative matrix factorization and pathway signature analyses. *Sci. Rep.* 8 (1), 9743. doi:10.1038/s41598-018-28066-w

Garrido, F., and Aptsiauri, N. (2019). Cancer immune escape: MHC expression in primary tumours versus metastases. *Immunology* 158 (4), 255–266. doi:10.1111/imm.13114

Gaynanova, I., and Li, G. (2019). Structural learning and integrative decomposition of multi-view data. *Biometrics* 75 (4), 1121–1132. doi:10.1111/biom.13108

Gonzalez, H., Hagerling, C., and Werb, Z. (2018). Roles of the immune system in cancer: From tumor initiation to metastatic progression. *Genes Dev.* 32 (19-20), 1267–1284. doi:10.1101/gad.314617.118

Gopalakrishnan, V., Helmink, B. A., Spencer, C. N., Reuben, A., and Wargo, J. A. (2018). The influence of the gut microbiome on cancer, immunity, and cancer immunotherapy. *Cancer Cell* 33 (4), 570–580. doi:10.1016/j.ccell.2018.03.015

Goswami, S., Chen, Y., Anandhan, S., Szabo, P. M., Basu, S., Blando, J. M., et al. (2020). *ARID1A* mutation plus CXCL13 expression act as combinatorial biomarkers to predict responses to immune checkpoint therapy in mUCC. *Sci. Transl. Med.* 12 (548), eabc4220. doi:10.1126/scitranslmed.abc4220

Gou, Q., Dong, C., Xu, H., Khan, B., Jin, J., Liu, Q., et al. (2020). PD-L1 degradation pathway and immunotherapy for cancer. *Cell Death Dis.* 11 (11), 955. doi:10.1038/s41419-020-03140-2

Griss, J., Bauer, W., Wagner, C., Simon, M., Chen, M., Grabmeier-Pfistershammer, K., et al. (2019). B cells sustain inflammation and predict response to immune checkpoint blockade in human melanoma. *Nat. Commun.* 10 (1), 4186. doi:10.1038/s41467-019-12160-2

Hamamoto, R., Takasawa, K., Machino, H., Kobayashi, K., Takahashi, S., Bolatkan, A., et al. (2022). Application of non-negative matrix factorization in oncology: One approach for establishing precision medicine. *Brief. Bioinform.* 23 (4), bbac246. doi:10.1093/bib/bbac246

He, K., Zhang, X., Ren, S., and Sun, J. (2016). *Deep residual learning for image recognition* in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet], Las Vegas, NV, USA, cité 10 janv 2022 (IEEE), 770–778. Disponible sur: http://ieeexplore.ieee.org/document/7780459/.

Helmink, B. A., Reddy, S. M., Gao, J., Zhang, S., Basar, R., Thakur, R., et al. (2020). B cells and tertiary lymphoid structures promote immunotherapy response. *Naturejanv* 23577 (7791), 549–555. doi:10.1038/s41586-019-1922-8

Herrmann, M., Probst, P., Hornung, R., Jurinovic, V., and Boulesteix, A. L. (2021). Large-scale benchmark study of survival prediction methods using multi-omics data. *Brief. Bioinform.* 22 (3), bbaa167. doi:10.1093/bib/bbaa167

Hiley, C., de Bruin, E. C., McGranahan, N., and Swanton, C. (2014). Deciphering intratumor heterogeneity and temporal acquisition of driver events to refine precision medicine. *Genome Biol.* 15 (8), 453. doi:10.1186/s13059-014-0453-8

Hoadley, K. A., Yau, C., Wolf, D. M., Cherniack, A. D., Tamborero, D., Ng, S., et al. (2014). Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* 158 (4), 929–944. doi:10.1016/j.cell.2014.06.049

Huang, G., Huang, G. B., Song, S., and You, K. (2015). Trends in extreme learning machines: A review. *Neural Netw.* 61, 32–48. doi:10.1016/j.neunet.2014.10.001

Huang, Z., Zhan, X., Xiang, S., Johnson, T. S., Helm, B., Yu, C. Y., et al. (2019). Salmon: Survival analysis learning with multi-omics neural networks on breast cancer. *Front. Genet.* 10, 166. doi:10.3389/fgene.2019.00166

Hudson, T. J., Anderson, W., Artez, A., Barker, A. D., Bell, C., Bernabé, R. R., et al. (2010). International network of cancer genome projects. *Nature* 464 (7291), 993–998. doi:10.1038/nature08987

Hui, K. M. (1989). Re-expression of major histocompatibility complex (MHC) class I molecules on malignant tumor cells and its effect on host-tumor interaction. *Bioessays* 11 (1), 22–26. doi:10.1002/bies.950110107

Iwai, Y., Ishida, M., Tanaka, Y., Okazaki, T., Honjo, T., and Minato, N. (2002). Involvement of PD-L1 on tumor cells in the escape from host immune system and tumor immunotherapy by PD-L1 blockade. *Proc. Natl. Acad. Sci. U. S. A.* 99 (19), 12293–12297. doi:10.1073/pnas.192461099

Jackson, L. A., Wang, S. P., Nazar-Stewart, V., Grayston, J. T., and Vaughan, T. L. (2022). *Association of Chlamydia pneumoniae immunoglobulin A seropositivity and risk of lung cancer*, 5.

Jamal-Hanjani, M., Hackshaw, A., Ngai, Y., Shaw, J., Dive, C., Quezada, S., et al. (2014). Tracking genomic cancer evolution for precision medicine: The lung TRACERx study. *PLoS Biol.* 12 (7), e1001906. doi:10.1371/journal.pbio.1001906

Janda, J. M., and Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *J. Clin. Microbiol.* 45 (9), 2761–2764. doi:10.1128/JCM.01228-07

Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., and Kluger, Y. (2018). DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med. Res. Methodol.* 18 (1), 24. doi:10.1186/s12874-018-0482-1

Khoruts, A., Dicksved, J., Jansson, J. K., and Sadowsky, M. J. (2010). Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent Clostridium difficile-associated diarrhea. *J. Clin. Gastroenterol.* 44 (5), 354–360. doi:10.1097/MCG.0b013e3181c87e02

Knox, S. S. (2010). From « omics » to complex disease: A systems biology approach to gene-environment interactions in cancer. *Cancer Cell Int.* 10 (1), 11. doi:10.1186/1475-2867-10-11

Kong, H. H., Oh, J., Deming, C., Conlan, S., Grice, E. A., Beatson, M. A., et al. (2012). Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. *Genome Res.* 22 (5), 850–859. doi:10.1101/gr.131029.111

Krassowski, M., Das, V., Sahu, S. K., and Misra, B. B. (2020). State of the field in multi-omics research: From computational needs to data mining and sharing. *Front. Genet.* 11, 610798. doi:10.3389/fgene.2020.610798

Kuenzi, B. M., Park, J., Fong, S. H., Sanchez, K. S., Lee, J., Kreisberg, J. F., et al. (2020). Predicting drug response and synergy using a deep learning model of human cancer cells. *Cancer Cell* 38 (5), 672–684. e6. doi:10.1016/j.ccell.2020.09.014

Lagoa, R., Marques-da-Silva, D., Diniz, M., Daglia, M., and Bishayee, A. (2020). Molecular mechanisms linking environmental toxicants to cancer development: Significance for protective interventions with polyphenols. *Seminars Cancer Biol.* 1, 1. doi:10.1016/j.semcancer.2020.02.002

Lakshmanan, V. K., Ojha, S., and Jung, Y. D. (2020). A modern era of personalized medicine in the diagnosis, prognosis, and treatment of prostate cancer. *Comput. Biol. Med.* 126, 104020. doi:10.1016/j.compbiomed.2020.104020

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521 (7553), 436–444. doi:10.1038/nature14539

Lee, D. D., and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature* 401 (6755), 788–791. doi:10.1038/44565

Li, P., and Chen, S. (2016). A review on Gaussian process latent variable models. *CAAI Trans. Intell. Technol.* 1 (4), 366–376. doi:10.1016/j.trit.2016.11.004

Li, T., Fu, J., Zeng, Z., Cohen, D., Li, J., Chen, Q., et al. (2020). TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* 48 (1), W509–W514. doi:10.1093/nar/gkaa407

Litchfield, K., Reading, J. L., Puttick, C., Thakkar, K., Abbosh, C., Bentham, R., et al. (2021). Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* 184 (3), 596–614. e14. doi:10.1016/j.cell.2021.01.002

Lock, E. F., Hoadley, K. A., Marron, J. S., and Nobel, A. B. (2013). Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. *Ann. Appl. Stat.* 7, 523–542. doi:10.1214/12-AOAS597

Ma, M., Chen, Y., Chong, X., Jiang, F., Gao, J., Shen, L., et al. (2021). Integrative analysis of genomic, epigenomic and transcriptomic data identified molecular subtypes of esophageal carcinoma. *Aging* 13 (5), 6999–7019. doi:10.18632/aging.202556

Ma, X., Liu, Y., Liu, Y., Alexandrov, L. B., Edmonson, M. N., Gawad, C., et al. (2018). Pan-cancer genome and transcriptome analyses of 1, 699 paediatric leukaemias and solid tumours. *Nature* 555 (7696), 371–376. doi:10.1038/nature25795

Mangul, S., Martin, L. S., Hill, B. L., Lam, A. K. M., Distler, M. G., Zelikovsky, A., et al. (2019). Systematic benchmarking of omics computational tools. *Nat. Commun.* 10 (1), 1393. doi:10.1038/s41467-019-09406-4

Marusyk, A., Almendro, V., and Polyak, K. (2012). Intra-tumour heterogeneity: A looking glass for cancer? *Nat. Rev. Cancer* 12 (5), 323–334. doi:10.1038/nrc3261

Marusyk, A., Janiszewska, M., and Polyak, K. (2020). Intratumor heterogeneity: The rosetta stone of therapy resistance. *Cancer Cell* 37 (4), 471–484. doi:10.1016/j.ccell.2020.03.007

Marx, V. (2020). Bench pressing with genomics benchmarkers. *Nat. Methods* 17 (3), 255–258. doi:10.1038/s41592-020-0768-1

McAllister, K., Mechanic, L. E., Amos, C., Aschard, H., Blair, I. A., Chatterjee, N., et al. (2017). Current challenges and new opportunities for gene-environment interaction studies of complex diseases. *Am. J. Epidemiol.* 186 (7), 753–761. doi:10.1093/aje/kwx227

Menyhárt, O., and Győrffy, B. (2021). Multi-omics approaches in cancer research with applications in tumor subtyping, prognosis, and diagnosis. *Comput. Struct. Biotechnol. J.* 19, 949–960. doi:10.1016/j.csbj.2021.01.009

Min, E. J., Chang, C., and Long, Q. (2018). "Generalized bayesian factor Analysis for integrative clustering with applications to multi-omics data," in 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA) [Internet], Turin, Italy, cité 16 août 2022 (IEEE), 109–119. Disponible sur: https://ieeexplore.ieee.org/document/8631499/.

Mirza, B., Wang, W., Wang, J., Choi, H., Chung, N. C., and Ping, P. (2019). Machine learning and integrative analysis of biomedical big data. *Genes* 10 (2), 87. doi:10.3390/genes10020087

Mo, Q., Li, R., Adeegbe, D. O., Peng, G., and Chan, K. S. (2020). Integrative multi-omics analysis of muscle-invasive bladder cancer identifies prognostic biomarkers for frontline chemotherapy and immunotherapy. *Commun. Biol.* 3 (1), 784. doi:10.1038/s42003-020-01491-2

Mo, Q., Shen, R., Guo, C., Vannucci, M., Chan, K. S., and Hilsenbeck, S. G. (2018). A fully Bayesian latent variable model for integrative clustering analysis of multi-type omics data. *Biostatistics* 19 (1), 71–86. doi:10.1093/biostatistics/kxx017

Mo, Q., Wang, S., Seshan, V. E., Olshen, A. B., Schultz, N., Sander, C., et al. (2013). Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proc. Natl. Acad. Sci. U. S. A.* 110 (11), 4245–4250. doi:10.1073/pnas.1208949110

Mobadersany, P., Yousefi, S., Amgad, M., Gutman, D. A., Barnholtz-Sloan, J. S., Velázquez Vega, J. E., et al. (2018). Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc. Natl. Acad. Sci. U. S. A.* 115 (13), E2970–E2979. doi:10.1073/pnas.1717139115

Needham, C. J., Bradford, J. R., Bulpitt, A. J., and Westhead, D. R. (2007). A primer on learning in Bayesian networks for computational biology. *PLoS Comput. Biol.* 3 (8), e129. doi:10.1371/journal.pcbi.0030129

Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* 37 (7), 773–782. doi:10.1038/s41587-019-0114-2

Nicora, G., Vitali, F., Dagliati, A., Geifman, N., and Bellazzi, R. (2020). Integrated multi-omics analyses in oncology: A review of machine learning methods and tools. *Front. Oncol.* 10, 1030. doi:10.3389/fonc.2020.01030

Nisar, S., Bhat, A. A., Hashem, S., Yadav, S. K., Rizwan, A., Singh, M., et al. (2020). Non-invasive biomarkers for monitoring the immunotherapeutic response to cancer. *J. Transl. Med.* 18 (1), 471. doi:10.1186/s12967-020-02656-7

Nishino, M., Ramaiya, N. H., Hatabu, H., and Hodi, F. S. (2017). Monitoring immune-checkpoint blockade: Response evaluation and biomarker development. *Nat. Rev. Clin. Oncol.* 14 (11), 655–668. doi:10.1038/nrclinonc.2017.88

Pei, A. Y., Oberdorf, W. E., Nossa, C. W., Agarwal, A., Chokshi, P., Gerz, E. A., et al. (2010). Diversity of 16S rRNA genes within individual prokaryotic genomes. *Appl. Environ. Microbiol.* 76 (12), 3886–3897. doi:10.1128/AEM.02953-09

Petalidis, L. P., Oulas, A., Backlund, M., Wayland, M. T., Liu, L., Plant, K., et al. (2008). Improved grading and survival prediction of human astrocytic brain tumors by artificial neural network analysis of gene expression microarray data. *Mol. Cancer Ther.* 7 (5), 1013–1024. doi:10.1158/1535-7163.MCT-07-0177

Pierre-Jean, M., Mauger, F., Deleuze, J. F., and Le Floch, E. P. IntM. F. (2022). PIntMF: Penalized integrative matrix factorization method for multi-omics data. *Bioinformatics* 38 (4), 900–907. doi:10.1093/bioinformatics/btab786

Poirion, O. B., Jing, Z., Chaudhary, K., Huang, S., and Garmire, L. X. (2021). DeepProg: An ensemble of deep-learning and machine-learning models for prognosis prediction using multi-omics data. *Genome Med.* 13 (1), 112. doi:10.1186/s13073-021-00930-x

Pucher, B. M., Zeleznik, O. A., and Thallinger, G. G. (2019). Comparison and evaluation of integrative methods for the analysis of multilevel omics data: A study based on simulated and experimental cancer data. *Brief. Bioinform.* 20 (2), 671–681. doi:10.1093/bib/bby027

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* 35 (9), 833–844. doi:10.1038/nbt.3935

Racle, J., and Gfeller, D. (2020). Epic: A tool to estimate the proportions of different cell types from bulk gene expression data. *Methods Mol. Biol.* 2120, 233–248. doi:10.1007/978-1-0716-0327-7_17

Rajbhandari, P., Lopez, G., Capdevila, C., Salvatori, B., Yu, J., Rodriguez-Barrueco, R., et al. (2018). Cross-cohort analysis identifies a TEAD4–MYCN positive feedback loop as the core regulatory element of high-risk neuroblastoma. *Cancer Discov.* 8 (5), 582–599. doi:10.1158/2159-8290.CD-16-0861

Rappoport, N., and Shamir, R. (2018). Multi-omic and multi-view clustering algorithms: Review and cancer benchmark. *Nucleic Acids Res.* 46 (20), 10546–10562. doi:10.1093/nar/gky889

Riaz, N., Havel, J. J., Makarov, V., Desrichard, A., Urba, W. J., Sims, J. S., et al. (2017). Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell* 171 (4), 934–949. e16. doi:10.1016/j.cell.2017.09.028

Riquelme, E., Zhang, Y., Zhang, L., Montiel, M., Zoltan, M., Dong, W., et al. (2019). Tumor microbiome diversity and composition influence pancreatic cancer outcomes. *Cell* 178 (4), 795–806. e12. doi:10.1016/j.cell.2019.07.008

Roh, V., Abramowski, P., Hiou-Feige, A., Cornils, K., Rivals, J. P., Zougman, A., et al. (2018). Cellular barcoding identifies clonal substitution as a hallmark of local recurrence in a surgical model of head and neck squamous cell carcinoma. *Cell Rep.* 25 (8), 2208–2222. e7. doi:10.1016/j.celrep.2018.10.090

Rosenfeld, A. M., Meng, W., et al. (2018). ImmuneDB, a novel tool for the analysis, storage, and dissemination of immune repertoire sequencing data. *Front. Immunol.* 9, 2107. doi:10.3389/fimmu.2018.02107

Routy, B., Le Chatelier, E., Derosa, L., Duong, C. P. M., Alou, M. T., Daillère, R., et al. (2018). Gut microbiome influences efficacy of PD-1–based immunotherapy against epithelial tumors. *Science* 359 (6371), 91–97. doi:10.1126/science.aan3706

Santucci, C., Carioli, G., Bertuccio, P., Malvezzi, M., Pastorino, U., Boffetta, P., et al. (2020). Progress in cancer mortality, incidence, and survival: A global overview. *Eur. J. Cancer Prev.* 29 (5), 367–381. doi:10.1097/CEJ.0000000000000594

Sathyanarayanan, A., Gupta, R., Thompson, E. W., Nyholt, D. R., Bauer, D. C., and Nagaraj, S. H. (2020). A comparative study of multi-omics integration tools for cancer driver gene identification and tumour subtyping. *Brief. Bioinform.* 21 (6), 1920–1936. doi:10.1093/bib/bbz121

Sepich-Poore, G. D., Zitvogel, L., Straussman, R., Hasty, J., Wargo, J. A., and Knight, R. (2021). The microbiome and human cancer. *Science* 371 (6536), eabc4552. doi:10.1126/science.abc4552

Sharifi-Noghabi, H., Zolotareva, O., Collins, C. C., and Ester, M. (2019). Moli: Multi-omics late integration with deep neural networks for drug response prediction. *Bioinformatics* 35 (14), i501–i509. doi:10.1093/bioinformatics/btz318

Shen, R., Olshen, A. B., and Ladanyi, M. (2009). Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics* 25 (22), 2906–2912. doi:10.1093/bioinformatics/btp543

Sherif, S., Roelands, J., Mifsud, W., Ahmed, E., Mifsud, B., Bedognetti, D., et al. (2021). The Immune landscape of pediatric solid tumors [Internet]. *Cancer Biol.* 1, 1. doi:10.1101/2021.05.04.442503

Simonyan, K., and Zisserman, A. (2015). *Very deep convolutional networks for large-scale image recognition.* arXiv:14091556 [cs] [Internet]. 10 avr 2015 [cité 10 janv 2022]; Disponible sur: http://arxiv.org/abs/1409.1556.

Singhal, N., Kumar, M., Kanaujia, P. K., and Virdi, J. S. (2015). MALDI-TOF mass spectrometry: An emerging technology for microbial identification and diagnosis. *Front. Microbiol. [Internet]* 5, 6. doi:10.3389/fmicb.2015.00791/abstract

Skoulidis, F., and Heymach, J. V. (2019). Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy. *Nat. Rev. Cancer* 19 (9), 495–509. doi:10.1038/s41568-019-0179-8

Song, M., Chan, A. T., and Sun, J. (2020). Influence of the gut microbiome, diet, and environment on risk of colorectal cancer. *Gastroenterology* 158 (2), 322–340. doi:10.1053/j.gastro.2019.06.048

Song, M., Greenbaum, J., Luttrell, J., Zhou, W., Wu, C., Shen, H., et al. (2020). A review of integrative imputation for multi-omics datasets. *Front. Genet.* 11, 570255. doi:10.3389/fgene.2020.570255

Song, Q., Zhu, X., Jin, L., Chen, M., Zhang, W., and Su, J. (2022). Smgr: A joint statistical method for integrative analysis of single-cell multi-omics data. *Nar. Genom. Bioinform.* 4 (3), lqac056. doi:10.1093/nargab/lqac056

Song, X., Ji, J., Gleason, K. J., Yang, F., Martignetti, J. A., Chen, L. S., et al. (2019). Insights into impact of DNA copy number alteration and methylation on the proteogenomic landscape of human ovarian cancer via a multi-omics integrative analysis. *Mol. Cell. Proteomics* 18 (8), S52–S65. doi:10.1074/mcp.RA118.001220

Spencer, M. D., Hamp, T. J., Reid, R. W., Fischer, L. M., Zeisel, S. H., and Fodor, A. A. (2011). Association between composition of the human gastrointestinal microbiome and development of fatty liver with choline deficiency. *Gastroenterology* 140 (3), 976–986. doi:10.1053/j.gastro.2010.11.049

Subramanian, I., Verma, S., Kumar, S., Jere, A., and Anamika, K. (2020). Multi-omics data integration, interpretation, and its application. *Bioinform. Biol. Insights* 14, 1. doi:10.1177/1177932219899051

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet], Boston, MA, USA, cité 10 janv 2022 (IEEE), 1–9. Disponible sur: http://ieeexplore.ieee.org/document/7298594/.

Tang, S., Ning, Q., Yang, L., Mo, Z., and Tang, S. (2020). Mechanisms of immune escape in the cancer immune cycle. *Int. Immunopharmacol.* 86, 106700. doi:10.1016/j.intimp.2020.106700

Tebani, A., Afonso, C., Marret, S., and Bekri, S. (2016). Omics-based strategies in precision medicine: Toward a paradigm shift in inborn errors of metabolism investigations. *Int. J. Mol. Sci.* 17 (9), 1555. doi:10.3390/ijms17091555

Thompson, J., Christensen, B., and Marsit, C. (2018). Pan-cancer analysis reveals differential susceptibility of bidirectional gene promoters to DNA methylation, somatic mutations, and copy number alterations. *Int. J. Mol. Sci.* 19 (8), 2296. doi:10.3390/ijms19082296

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T. H., et al. (2018). The immune landscape of cancer. *Immunity* 48 (4), 812–830. e14. doi:10.1016/j.immuni.2018.03.023

Tin Kam, H. (1995). "Random decision forests," in Proceedings of 3rd International Conference on Document Analysis and Recognition [Internet], Montreal, Que., Canada, cité 10 janv 2022 (IEEE Comput. Soc. Press), 278–282. Disponible sur: http://ieeexplore.ieee.org/document/598994/.

Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). The cancer genome atlas (TCGA): An immeasurable source of knowledge. *Contemp. Oncol.* 1A, 68–77. doi:10.5114/wo.2014.47136

Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R., and Gordon, J. I. (2009). The effect of diet on the human gut microbiome: A metagenomic analysis in humanized gnotobiotic mice. *Sci. Transl. Med. [Internet]* 1 (6), 1. doi:10.1126/scitranslmed.3000322

van Ijzendoorn, D. G. P., Szuhai, K., Briaire-de Bruijn, I. H., Kostine, M., Kuijjer, M. L., and Bovée, J. V. M. G. (2019). Machine learning analysis of gene expression data reveals novel diagnostic and prognostic biomarkers and identifies therapeutic targets for soft tissue sarcomas. *PLoS Comput. Biol.* 15 (2), e1006826. doi:10.1371/journal.pcbi.1006826

Wang, J., and éditeur (2009). *Encyclopedia of data warehousing and mining.* Second Edition. Hershey, Pennsylvania: IGI Global. [Internet]cité 16 août 2022]. Disponible sur: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-60566-010-3.

Weber, L. M. (2019). *Essential guidelines for computational method benchmarking*, 12.

Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., et al. (2013). The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* 45 (10), 1113–1120. doi:10.1038/ng.2764

Weitschek, E., Felici, G., and Bertolazzi, P. (2012). "Mala: A microarray clustering and classification software," in 2012 23rd International Workshop on Database and Expert Systems Applications [Internet], Vienna, Austria, cité 5 janv 2022 (IEEE), 201–205. Disponible sur: http://ieeexplore.ieee.org/document/6327426/.

Witkowski, M. T., Dolgalev, I, Evensen, N. A., Ma, C., Chambers, T., Roberts, K. G., et al. (2020). Extensive remodeling of the immune microenvironment in B cell acute lymphoblastic leukemia. *Cancer Cell* 37 (6), 867–882. e12. doi:10.1016/j.ccell.2020.04.015

Wong, M., Mayoh, C., Lau, L. M. S., Khuong-Quang, D. A., Pinese, M., Kumar, A., et al. (2020). Whole genome, transcriptome and methylome profiling enhances actionable target discovery in high-risk pediatric cancer. *Nat. Med.* 26 (11), 1742–1753. doi:10.1038/s41591-020-1072-4

Wörheide, M. A., Krumsiek, J., Kastenmüller, G., and Arnold, M. (2021). Multi-omics integration in biomedical research – a metabolomics-centric review. *Anal. Chim. Acta* 1141, 144–162. doi:10.1016/j.aca.2020.10.038

Wu, B., Lu, X., Shen, H., Yuan, X., Wang, X., Yin, N., et al. (2020). Intratumoral heterogeneity and genetic characteristics of prostate cancer. *Int. J. Cancer* 146 (12), 3369–3378. doi:10.1002/ijc.32961

Yang, Z., and Michailidis, G. (2015). A non-negative matrix factorization method for detecting modules in heterogeneous omics multi-modal data. *Bioinformatics* 32, 1–8. doi:10.1093/bioinformatics/btv544

Yuan, L., Zhao, J., Sun, T., and Shen, Z. (2021). A machine learning framework that integrates multi-omics data predicts cancer-related LncRNAs. *BMC Bioinforma.* 22 (1), 332. doi:10.1186/s12859-021-04256-8

Zeng, D., Ye, Z., Shen, R., Yu, G., Wu, J., Xiong, Y., et al. (2021). Iobr: Multi-Omics immuno-oncology biological research to decode tumor microenvironment and signatures. *Front. Immunol.* 12, 687975. doi:10.3389/fimmu.2021.687975

Zhang, S., Liu, C. C., Li, W., Shen, H., Laird, P. W., and Zhou, X. J. (2012). Discovery of multi-dimensional modules by integrative analysis of cancer genomic data. *Nucleic Acids Res.* 40 (19), 9379–9391. doi:10.1093/nar/gks725

Zhao, Y., Gao, Y., Xu, X., Zhou, J., and Wang, H. (2021). Multi-omics analysis of genomics, epigenomics and transcriptomics for molecular subtypes and core genes for lung adenocarcinoma. *BMC Cancer* 21 (1), 257. doi:10.1186/s12885-021-07888-4

Zheng, L., Qin, S., Si, W., Wang, A., Xing, B., Gao, R., et al. (2021). Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science* 374 (6574), abe6474. doi:10.1126/science.abe6474

Zhou, J. X., Taramelli, R., Pedrini, E., Knijnenburg, T., and Huang, S. (2017). Extracting intercellular signaling network of cancer tissues using ligand-receptor expression patterns from whole-tumor and single-cell transcriptomes. *Sci. Rep.* 7 (1), 8815. doi:10.1038/s41598-017-09307-w

Zhu, W., Xie, L., Han, J., and Guo, X. (2020). The application of deep learning in cancer prognosis prediction. *Cancers* 12 (3), 603. doi:10.3390/cancers12030603