EPMA Journal (2017) 8:51–60
DOI 10.1007/s13167-017-0083-9

REVIEW

Pattern recognition for predictive, preventive, and personalized medicine in cancer

Tingting Cheng[1,2,3] · Xianquan Zhan[1,2,3,4]

Received: 31 December 2016 / Accepted: 5 February 2017 / Published online: 9 March 2017

Abstract Predictive, preventive, and personalized medicine (PPPM) is the hot spot and future direction in the field of cancer. Cancer is a complex, whole-body disease that involved multi-factors, multi-processes, and multi-consequences. A series of molecular alterations at different levels of genes (genome), RNAs (transcriptome), proteins (proteome), peptides (peptidome), metabolites (metabolome), and imaging characteristics (radiome) that resulted from exogenous and endogenous carcinogens are involved in tumorigenesis and mutually associate and function in a network system, thus determines the difficulty in the use of a single molecule as biomarker for personalized prediction, prevention, diagnosis, and treatment for cancer. A key molecule-panel is necessary for accurate PPPM practice. Pattern recognition is an effective methodology to discover key molecule-panel for cancer. The modern omics, computation biology, and systems biology technologies lead to the possibility in recognizing really reliable molecular pattern for PPPM practice in cancer. The present article reviewed the pathophysiological basis, methodology, and perspective usages of pattern recognition for PPPM in cancer so that our previous opinion on multi-parameter strategies for PPPM in cancer is translated into real research and development of PPPM or precision medicine (PM) in cancer.

Keywords Pattern recognition · Predictive preventive personalized medicine · Genomics · Transcriptomics · Proteomics · Peptidomics · Metabolomics · Radiomics · Systems biology

Abbreviations

ADTEx	Aberration detection in tumour exome
CTC	Circulating tumor cell
ctDNA	Circulating tumor DNA
DF-SNPs	Decision forest for SNPs
LMIs	Low-mass ions
MALDI-TOF-MS	Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry
mRNA	Messenger RNA
ncRNA	Non-coding RNA
PM	Precision medicine
PPPM	Predictive, preventive and personalized medicine
RT-PCR	Real-time quantitative PCR
SNP	Single nucleotide polymorphisms
VOC	Volatile organic compounds

Introduction

In the last decades, the incidence of cancer rose year by year, a number of people die of it, and cancer is the biggest threat to human health. A growing number of studies confirm that

✉ Xianquan Zhan
yjzhan2011@gmail.com

[1] Key Laboratory of Cancer Proteomics of Chinese Ministry of Health, Xiangya Hospital, Central South University, 87 Xiangya Road, Changsha, Hunan 410008, People's Republic of China

[2] Hunan Engineering Laboratory for Structural Biology and Drug Design, Xiangya Hospital, Central South University, 87 Xiangya Road, Changsha, Hunan 410008, People's Republic of China

[3] State Local Joint Engineering Laboratory for Anticancer Drugs, Xiangya Hospital, Central South University, 87 Xiangya Road, Changsha, Hunan 410008, People's Republic of China

[4] The State Key Laboratory of Medical Genetics, Central South University, 88 Xiangya Road, Changsha, Hunan 410008, People's Republic of China

tumor is a chronic disease involving the whole body. The growth of tumors is involved in many stages and complex processes, and in many genes and molecular events including multi-gene mutations, such as activation of oncogenes and inactivation of tumor suppressor genes. In the present documented literature, most of them endeavor on the effect of single factor on the development of cancer in the hypothetical conditions. However, some studies found that not only one molecular event leads to the occurrence of cancer. A typical cancer occurrence model needs the mutation of two to eight driver genes [1]. The mutation of passenger genes is not able to lead the development of cancer [2]. The goal of studies should focus on a panel of gene mutations, which is called gene pattern mutation. Depending on the central dogma, gene pattern mutation may affect a series of mRNA and protein expressions. In order to set diagnosis models based on differentially expressed proteins or peptides between tumor tissues and normal tissues, this pattern would avoid the result of low sensitivity of a single-tumor marker or low specificity of a large number of samples. In addition, with the development of cancer biomarkers, one found that the change of key molecule panel in gene and protein sequences initiates the tumoregenesis. As different individual has different key molecule panel, clinical doctors can use different targeted drugs to prevent the occurrence of tumor. The multi-parameter systematic strategy for predictive, preventive, and personalized medicine (PPPM) in cancer was initially conceived by the Zhan and Desiderio [3]. Moreover, cancer biology has gradually shifted to the era of precision cancer medicine [4]. The focus of this review article is on the use of tumor biological characteristics changes to guide the patient's diagnosis, treatment, and prognosis judgment.

Recently, more and more patients are putting attentions on precision therapy, which needs more and more biomarkers to be found. The best optimal biomarker is only changes in cancer patients and can be easily detected. By far the most common cancer biomarkers are generally to detect the removed cancer tissues, which is an invasive operation. If the tumor is too little to be found, or it is difficult to get the tumor tissue, those biomarkers are helpless.

As we know, the growth of cancer is a complicate progress. From DNA, RNA, protein to metabolite, all the differences in the levels of DNA, RNA, protein, and metabolite between cancer patients and health persons could be called biomarkers. Although many biomarkers have been found, but less invasive, early and effective biomarkers are still limited. Nowadays, the common biomarkers that are used in clinic are always from the four ways: (i) metabolic products of tumor cells, (ii) abnormal differentiation of cellular gene products, (iii) tumor necrosis and exfoliation of tumor cells release into the blood circulation, and (iv) cell reactive products of tumor host cells. However, all of these biomarkers only can be detected when cancer occurred. Before cancer occurred, DNA/RNA/protein and the environment changes in normal cells could make normal cell changes into differentiation disorder cell, which is considered the cause of cancer. With the development of image technology, it was founded that imaging features of cancer appearance have a close relationship with the diagnosis and prognosis of patients. Imaging features could become a new type of biomarkers. In terms of function, biomarkers can be divided into two categories: (i) contribution to the mechanism and therapeutic targets, and (ii) contribution to prediction, diagnostic test, and prognosis assessment. The first kind of biomarker has a causal relationship with the occurrence and development of disease, which can directly address the pathogenesis of the disease. It is generally the key sites in the cell signal pathways, such as P53 in nasopharyngeal carcinoma (NPC) [5]. The second type of biomarker may have no causal relationship with the occurrence and development of the disease; however, they should not only have specificity but also achieve a certain amount of change to be easily detected. Not all biomarkers need to be changed before the disease occurs, only with the detection of the type of biomarkers related needs, another type does not need (Fig. 1). In this review article, pattern recognition exactly means to recognize pattern biomarker, namely to use a set of patterns that is composed of several biomarkers to improve the accuracy and specificity of prediction, diagnosis, prognosis, and prevention/therapy of tumor.

Pathophysiological basis of pattern recognition for PPPM in cancer

Human displays the most complicated and diverse phenotypic traits relative to any other living organisms [6]. The earlier studies predict that only 0.1% of the entire genome differs between individuals. Those genomic diversities are affected by ethnic and geographic differences in a wide variety of traits [7]. The high penetrance of heritable mutations and subtle variants contribute to somatic alterations. All of those lead to cellular traits that facilitate carcinogenesis, which determines individual's risk to develop certain cancers [8]. Cancer biomarkers play important roles in proliferation, invasion, and metastasis, and are related to prevention, diagnosis, and treatment including acquired drug resistance. Therefore, in modern oncology, the most important goal is to find the ways to effectively control tumor heterogeneity and translate these achievements to benefit patients.

Up to date, clinical trial allocation has been based on the right target, right drug, and right moment, so most trials focus on those patients who share the similar targetable biomarkers. However, cells within tumors have diverse genomes and epigenomes, and interact differentially with their surrounding microenvironment that includes extracellular matrix, inflammatory cells, immune cells, endothelial cells, fibroblasts, etc.

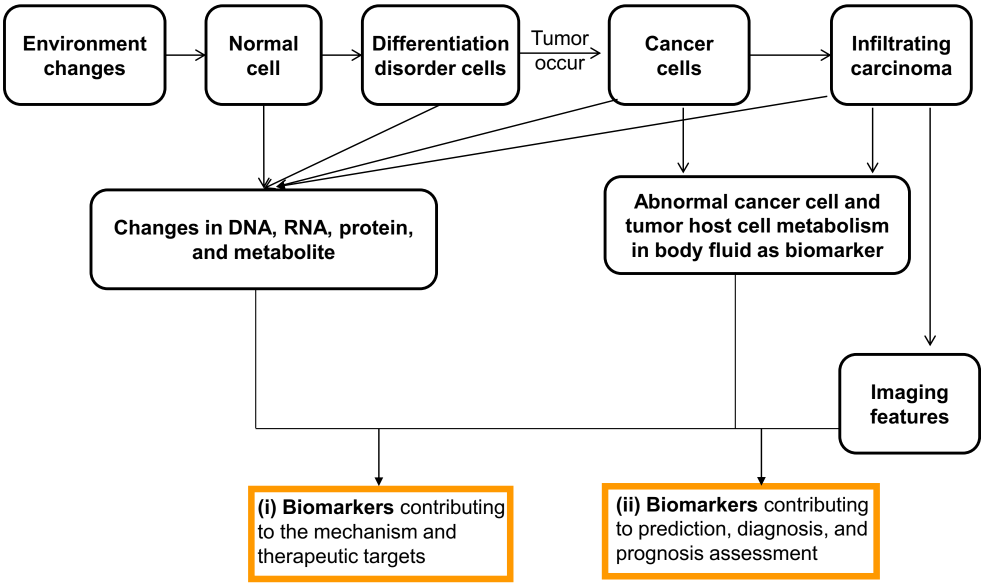
Environment changes
Normal cell
Differentiation disorder cells
Tumor occur
Cancer cells
Infiltrating carcinoma
Changes in DNA, RNA, protein, and metabolite
Abnormal cancer cell and tumor host cell metabolism in body fluid as biomarker
Imaging features
(i) Biomarkers contributing to the mechanism and therapeutic targets
(ii) Biomarkers contributing to prediction, diagnosis, and prognosis assessment

Fig. 1 Types of biomarkers for cancer

All those factors generate intra-tumor heterogeneity, which has critical implications for treating cancer patients. Tumor diversity poses a challenge for managing the treatment of cancer patients [9]. Clinical trial allocation would be based not only on the characterization of tumor biomarkers but also pay more attentions on tumor heterogeneity.

Although whole-genome, whole-exome, and whole-transcriptome sequencing offer an appropriate approach and opportunities for discovery, their immediate effect on clinical decision-making is limited, as only a fraction of cancer genes are well characterized in terms of biology and therapeutic relevance. In modern oncology, the most important goal is to find the ways to effectively control tumor heterogeneity and translate these achievements to benefit patients. Because the development of cancer is a complicated process and affected by many factors, therefore a single biomarker that resolves the relative problems of a cancer is a false appearance [10]. As we had mentioned above, more than one key locus changes lead the occurrence of tumor, the most suitable way is to find the core parameters for the specific trials and make those parameters into a pattern.

Based on Baye's Rule, if a novel biomarker (or a combination of biomarkers) diagnosis assay is 95% effective in detecting a certain disease and 0.5% of the population has the disease, the probability that a person with a positive test result actually has the disease is only 32.3%. So when we use one biomarker, the positive rate is too low to predict disease. The positive rate can be improved with multiple independent diagnosis assays. For example, biomarkers A, B, and C have 32.3% detection probabilities, respectively. Then, the probability that a person with positive results from all three assays has the disease will be 68.97% [11]. Thereby, improving of detecting real positives needs more than one biomarker. Three or more biomarkers, which are related to tumors, can form one pattern in order to enhance the accuracy of cancer diagnosis.

Methodology of pattern recognition for PPPM in cancer

More endeavors could be put on finding less invasive, early and effective method of cancer diagnosis. According to the central dogma, genetic changes affect the RNA, then lead the changes of proteins. Those proteins directly act on the cells and result in the occurrence of cancer. In order to better predict the occurrence and progression of tumors, this section illustrates the new method in genome, transcriptome, proteome, metabolome, and radiome. The integrative pattern derived from biological omics data (genomics, transcriptomics, proteomics, metabolomics, and radiomics) with the development of new algorithm will effectively contribute to cancer precise medicine (Fig. 2).

Genomics

The entire human genome contains about 2.91 Gbp and more than 39,000 genes [12]. With the development of gene sequencing technology, the first generation of sequencing technology is gradually replaced by the second generation sequencing technology, the sequencing efficiency has been significantly improved, and the cost is lower than before, which provides technical support for large-scale sequencing.

Development of genomics and proteomics in cancer offers the possibility of molecular diagnostics in the levels of gene and protein. Genomic instability in cancer leads to abnormal genome copy number alterations (CNA) that are associated

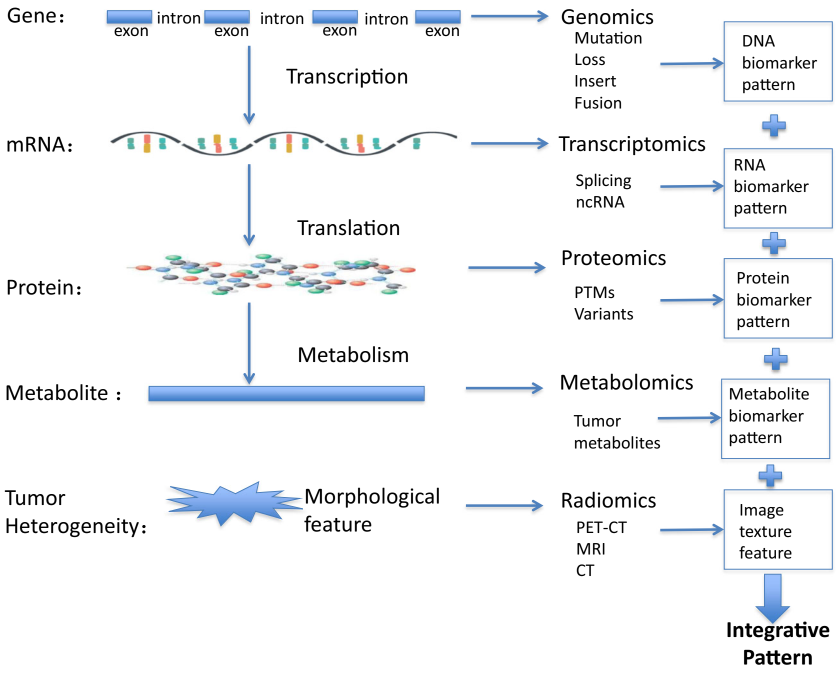

Fig. 2 Integrative pattern based on omics data

with the development and behavior of tumors. Large numbers of polymorphic CNA have been founded in the human genome [13]. Regional CNA has been demonstrated in tumors and linked to leading them to develop aggressive behavior. With the studies of gene expression patterns among different cell types (normal, pre-cancerous, and cancerous, and different types and stages), molecular diagnostics aimed to expose the "molecular signatures" to indicate those modes of peculiar pathology [14]. DNA microarrays, also known as "gene chip" or "DNA chip", have been successful because they allow researchers to monitor tens of thousands of one-time expression and hundreds of thousands genes. Single nucleotide polymorphisms (SNPs) are the most prevalent form of DNA variations in the human genome occurring about once per 100 to 300 bases [15]. SNPs contain insert, loss, and fusion. Many experiments confirmed that SNP could affect metabolism-related key enzyme activities, thereby affecting the efficacy of the tumor progression and drugs. It had examined the association between esophageal cancer risk and patterns of 61 SNPs in a case-control study for a population who has among the highest rates of esophageal squamous cell carcinoma. Another example, UGT1A1, is a very important gene in the prediction and therapy of cancer. UGT1A1*28, a relatively common gene variant of UGT1A1, is currently an extensively studied site in many different tumors such as colon cancer and leukemia [16]. UGT1A1*28 gene polymorphism refers to a TATA box with thymine adenine (TA) repeats [17]; for example, homozygous genotype TA6/6 refers to two wild-type gene (TA repeated six times) individuals; TA7/7 homozygous genotype that is TA7/7 refers to two UGT1A1*28 allele (seven TA repeats) individual. It cannot predict the prognosis and drug toxicity alone, but when it combines with UGT1A1*6 and MTHFR [18], they work.

A new method named Decision Forest for SNPs (DF-SNPs) has been developed from a novel adaptation of the Decision Forest pattern recognition. The DF-SNPs method can be used to differentiate esophageal squamous cell carcinoma cases from controls based on individual SNPs, SNP types, and SNP patterns [19]. However, with further research, scientists have found a SNP or simply CNA does not affect the overall development of the individual process of tumor. The occurrence of cancer is not simply a site change, but the change at multiple sites, so now gradually moving to study the composition of several mutation gene patterns. Those gene patterns may be related to one pathway, which is very important in the occurrence of cancer, or may be act synergistically in a key point. It has been found that unique pattern of component gene disruption in the NRF2 inhibitor KEAP1/CUL3/RBX1 E3-ubiquitin ligase complex in serous ovarian cancer. The KEAP1/CUL3/RBX1 E3-ubiquitin ligase complex is a regulator of NRF2 levels that is critical to initiate responses to oxidative stress [20].

Those methods described above depend on finding the key locus of its regulatory sites and longer study period. Biclustering techniques have become very popular in cancer genetics studies, which are expected to connect phenotypes to genotypes; for example, to identify subgroups of cancer patients based on the fact that they share similar gene expression patterns as well as to identify subgroups of genes that is specific to these subtypes of cancer, and therefore could serve as biomarkers [21].

Nowadays, another new way to get DNA information about tumor tissue is the circulating tumor cell (CTC), which is a general designation of all kinds of tumor cells in peripheral blood [22]. Compared to tumor tissue samples, blood samples were more easily acquired, less invasive, and can be repeatedly collected. It is an ideal source of specimens in clinical practice, which greatly improves the value of this method. Circulating tumor DNA (ctDNA) refers to the body of tumor cells by apoptosis after shedding or when released into the circulatory system; with the rapid development of gene sequencing, at present, one has been able to detect and count on it in the blood [23]. So ctDNAs are new types of biomarkers, which can be found mutation of key sites. In recent years, liquid biopsy based on ctDNA analysis has made great contribution to the molecular diagnosis and monitoring of cancer. With the developed in technique, BEAMing (beads, emulsion, amplification, and magnetics) [24] and CAPP-seq (cancer personalized profiling by deep sequencing) [25] were found to quantify ctDNA in blood. However, there have many mysteries about ctDNA, such as its size, existing form, mechanisms about released into blood stream, and its degradation rate in blood [26].

Transcriptomics

Recent progress in sequencing technology has significantly improved the ability of the researchers to study the nucleic acid level of biology. In the past years, scientists have put a lot of efforts to study the messenger RNAs (mRNAs), which carry genetic information, as a template when mRNA guides the protein synthesis. When the gene sequence of mRNA changes, the amino acid sequence of the protein will be correspondingly changed. Through these new powerful techniques, especially research in the field of noncoding RNA (ncRNA), ncRNA elements including multiple new and unique species were found and characterized. The current categories of ncRNAs include tRNA, rRNA, snoRNA, snRNA, piRNA, miRNA, and lncRNA [27].

MicroRNAs are a class of small ncRNAs with a sequence of approximately 21 bp that play a central role in the regulation of mRNA expression [28–32]. The discovery that microRNA expression is frequently dysregulated in a cancer-specific manner provides an opportunity to develop these RNAs as biomarkers for cancer detection [33–39]. However, because tumor-derived microRNAs can be present in blood and appear to be stable to certain degree and protect from endogenous ribonuclease activity in circulation, some studies have shown diagnostic and prognostic potential for circulating microRNAs [40–52].

The potential of circulating microRNAs as biomarkers for cancer early detection is particularly relevant to breast cancer that is the most common cancer in women, regardless of race or ethnicity, despite improvement in cancer screening and treatment strategies. In addition to cancers, circulating microRNAs, especially inflammation-related circulating microRNAs, may also be used as biomarkers for aging and other aging-related diseases [53, 54].

In traditionally, the expression levels of microRNAs were confirmed with a Taqman-based real-time quantitative PCR (RT-qPCR) using individual microRNA-specific primers and probes. It has been demonstrated that both miR-148b and miR-133a have potential to use as biomarkers for breast cancer detection. Moreover, the discovery of the role of miRNA in drug resistance and miR-polymorphisms to predict drug response has led to the development of a new field in biomedical science called miRNA pharmacogenomics, a study of the miRNAs and miR-polymorphisms affecting expressions of drug target genes, to predict drug behavior and to improve drug efficacy [55]. Several miRNAs were found to be associated (miR-192, miR-215, miR-140, miR-129, let-7, miR-181b, and miR-200) with chemoresistance by regulating key cell death pathways such as apoptosis and autophagy [56, 57]. The signature can be validated on a formalin-fixed paraffin-embedded (FFPE) tissue-specific and RT-PCR-based assay. The gene signature was further validated in an FFPE tissue cohort of 222 cases of primary clear cell renal cell carcinoma (ccRCC), with an overall sensitivity and specificity of 70% and 76%, respectively. The sensitivity was 59% and specificity was 74% for predicting metastasis from stage II patients. When it was used to predict for stage III patients, they were 80 and 83%, respectively. The signature was associated with the patient's cancer-specific survival and can be utilized as a predictive biomarker [58].

The largest group of ncRNAs are the long noncoding RNAs (lncRNAs) that perform a diverse set of functions within the cell. Importantly, lncRNAs have recently been implicated in the pathogenesis of multiple types of cancers, including breast, lung, gastric, liver, and prostate cancers [59]. The biological role of lncRNAs is still incompletely understood, but they have already been found to be prolific regulators of numerous cell processes. Some lncRNAs overlap with gene promoters and thus, transcription of these lncRNAs can interfere with nucleosome-deleted regions and histone modifications of nucleosomes in those promoters [60, 61]. Many lncRNAs have been confirmed to play important roles in cancers. Some have been implicated in a variety of cancers from different types of tissues, such as H19 and HOTAIR. H19 was among the earliest lncRNAs to be identified and it was touted as a potent tumor suppressor at the same time [62]. In the case of prostate cancer, four lncRNAs (PCAT-1, PCAT-5, MALAT1, and NEAT1) have been found to enhance these processes. Whereas, PCAT29 and DRAIC have been associated with inhibition of tumor growth [27, 63]. However, detection of lncRNA is easily affected by anticoagulant such as EDTA, and lncRNA is easily degraded by be other materials in the blood, so it cannot be long-term preserved. More studies are needed to solve these problems in the future.

Proteomics

Proteins directly regulate the growth and metabolism of cells in the human body, regulation of the protein alteration of key sites might inhibit the occurrence and growth of tumor. This is the theoretical basis of many chemotherapeutic drugs at present. In the last decade, the number of publications based on proteomics has dramatically increased. However, proteome is more complex than we had been imaged especially in behavior and structure. A single protein could be found different variants especially those variants have different functions in cells. Those variants from one protein are called as protein species or proteoform that has been defined at the chemical, molecular level [64]. Those protein species coded by the same gene are mainly derived from splicing and post-translational modifications (PTMs) [65–67]. It has reported that the ESAT-6 gene product of mycobacterium tuberculosis differentiates into at least eight protein species [68]. Furthermore, the environment can also affect the protein species, such as temperature or oxidative stress reaction.

The most commonly used methods to identify PTMs and protein species are 2D gel electrophoresis and mass spectrometry analysis. Studies show that the same protein was found at several different spots on 2D electrophoresis gels, and one 2D electrophoresis gel spot usually contains more than two proteins [64, 69]. In the last decade, imaging mass spectrometry has been incredible technological advances in its applications to biological samples. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) technology is widely applied in proteomics, such as serology tumor marker studies. Those identified proteins from cell, tissue, or the body provide a set of protein function and mode of information, to reflect the intracellular genetic characteristics and the effect of external factors. For example, glioma is a commom malignant brain tumor. The methods to diagnose glioma include CT and MRI. However, the misdiagnosis rate is still very high [70]. Some studies found that 11 peptide recognition and specific peak intensities are useful to diagnose glioma and its grading [71]. Therefore, identification of a large number of proteins from a biological specimen is a more overall detection in cancer studies.

Currently tumor tissue pathology are still the gold standard for diagnosis of tumors. So protein analysis of tumor tissue is more likely to be accepted. Mass spectrometry imaging (MSI) of biological tissue can provide topographic localization of biochemical information to complement the traditional pathology classification system. Among the MSI techniques currently available, three most commonly used techniques are MALDI, secondary ion mass spectrometry (SIMS), and desorption electrospray ionization (DESI). Until now, the routine clinical application of MSI approaches has been restricted by inherent time/cost demands and associated heavy analytical workload.

MSI offers a way to chemically map the tumor microenvironment intact, avoiding the need for time-consuming and disruptive procedural steps such as laser-capture microdissection. The inherently multidimensional nature of MSI datasets challenges conventional data processing method, but now the full potential of this emerging technique is still unfulfilled. Analysis results show that integration of MSI data and gene expression data is able to provide a meaningful discrimination between samples. Therefore, it is a useful tool in identity of large scale of potential biological information, such as between cancer patients and health people [72].

In the colorectal cancer tissue, unique lipid patterns were observed with MSI according to tissue type. A tissue recognition system using multivariate molecular ion patterns allowed highly accurate (>98%) identification of pixels according to morphology (cancer, healthy mucosa, smooth muscle, and microvasculature) [73].

Metabolomics

It has recently become clear that altered metabolic homeostasis plays important roles in carcinogenesis. Metabolism is directly or indirectly involved in every aspect of cellular functions. Metabolites commonly exist in the expired gas, tears, urine, saliva, CSF, and blood. Tumor-related metabolites can also be used as tumor biomarkers. Metabolomics was thought to reflect the status of any cell. Blood metabolites could be detected as low-mass ions (LMIs) by MS. A LMI discriminant equation (LOME) is constructed to investigate whether systematic LMI profiling might be applied to cancer screening. Colorectal cancer LOME demonstrated excellent discriminating power in a validation set with sensitivity/specificity of 93.21%/96.47%. Furthermore, in a fecal occult blood test (FOBT) of available validation samples, the discriminating power of CRC LOME was much stronger with sensitivity/specificity of 94.79%/97.96% than that of the FOBT with sensitivity/specificity of 50.00%/100.0%, which is the standard CRC screening tool [74].

The metabolism of tumor tissues in our body may produce some proteins or peptides that are different from normal tissues. Due to the rapid development of proteomic techniques, magnetic beads (liquid chip)-based MALDI-TOF-MS technology is used to screen distinctive biomarkers for lung adenocarcinoma (adCA) and to establish the diagnostic protein profiles. The profile gained by pattern recognition genetic algorithm that could distinguish adCAs from benign lung diseases was comprised of 4053.88, 4209.57, and 3883.33 Da with sensitivity of 80%, and specificity of 93%, while that could separate adCA from healthy control was comprised of 2951.83 and 4209.73 Da with sensitivity of 94%, and specificity of 95% [75]. Now many targeted therapies are used in cancer patient, which is based on several specific metabolic features of cancer cells.

There is a high demand for a simple and non-invasive test for selecting the individuals at increased risk. Over the past two decades, the analysis of volatile organic compounds (VOCs) has witnessed an enormous boost, as they have been described as a possible method to diagnose rapidly a variety of diseases, for example, cancers of the lung, breast, colon, prostate, liver, head-and-neck, as well as kidney disease, multiple sclerosis, and Parkinson's disease [76]. Predictive models were built employing discriminant factor analysis (DFA) pattern recognition, and their stability against possible confounding factors was tested. Complementary chemical analysis of the breath samples was performed using gas chromatography coupled with mass spectrometry [19].

Moreover, integrative approaches used to analyze the exhaled breath have demonstrated high sensitivity and specificity of this method for lung cancer diagnosis. Such integrative approaches include detection of breathprint by electronic nose or integrated analysis of wide range of VOCs detected by gas chromatography/mass spectrometry or related methods [77–79]. Apart from VOCs, tumor cells produce wide range of cytokines like IL-4, IL-6, IL-11, IL-15, TNF-a, TGF-b, and others, which activate body's immune system and change the metabolism of wide range of body cells [80, 81].

Evidently during the process of carcinogenesis, some longstanding changes develop also outside the tumor. These changes may be of immunological or genetic origin, based on observations that VOC pattern did not differ between the tumor stages. Applicable for such a purpose is electronic nose. Diagnostics using this device is simple, sufficiently accurate, inexpensive and noninvasive, allows online diagnosis, and can differentiate heterogeneous disorders. The information provided by this technique is not based on detecting single and separate molecular signals, but is exclusively derived from pattern recognition among an array of signals by using powerful bioinformatics [82]. Electronic nose is an instrument made up of different kind of chemical sensors combined with a pattern recognition system. The measurement in electronic nose is based on the different mechanisms—electrical resistance, ion gas, or colorimetric sensor response that differs regarding VOC molecular pattern [83].

Radiomics

Radiomics refers to the extraction and analysis of large amounts of advanced quantitative imaging features with high throughput from medical images obtained with computed tomography (CT), positron emission tomography (PET), or magnetic resonance imaging (MRI) [84]. It is proposed to reveal quantitatively predictive or prognostic associations between images and medical outcomes with analysis and mining of image feature data. The radiomics is a new field, which depends on the developed computer technology and advanced statistical methods. It may change many algorithms of region of interest (ROI) of the image data into high-resolution data mining of characteristics [85]. Through high-throughput quantitative analysis of digital image data, various target information obtain high fidelity phenotypic evaluation of tumor (phenotypes), including various levels of morphology, molecules, and genes [86].

Radiomics has great potential to guide cancer treatment, prognosis, and curative effect evaluation, because it can provide insight into the evaluation of the tumor completely, and can reflect the tumor development, progression, and response to therapy. Compared to the traditional methods of molecular biology, radiomics has the advantages of complete information and good repeatability, and is non-invasive, convenient, and cheap. In recent years, the study of prediction model of clinical efficacy or side effects based on the imaging features and molecular markers is more concentrated in the analysis of MRI, CT, and PET-CT image features. Scientists use MRI images to predict the effect of NPC radiotherapy and chemotherapy. The results showed that the texture features extracted from T1, T2, and DWI images can be used as the prediction index of NPC radiotherapy and chemotherapy. It is worth mentioning that the accuracy of T1 images is the highest, up to 95.2% [3]. In the next step, we can construct prediction model of clinical efficacy or side effects by pattern that integrates the imaging features and molecular markers in order to increase specificity and sensitivity.

New algorithm

The differences between cancer patients and health persons contain varied genes and proteins. However, how to find out those proteins or genes, which has statistically significant difference, is still a big problem. With the development of bioinformatics, a lot of large biological information database is established. In the past decade, complex networks have been widely used to analyze complex systems and they were proposed as a new tool to analyze the spectra extracted from biological samples. Three customary feature selection algorithms have been presented, including the binning of spectral data and the use of information theory metrics. Such algorithms are compared by assessing the score obtained in a classification task, where healthy subjects and people suffering from different types of cancers should be discriminated. Results show that mutual information outperforms the more classical data binning [87]. A new method that is combined into a package named ADTEx (Aberration Detection in

Tumour Exome) was established to infer copy number and genotypes using whole exome data from paired tumor/normal samples. ADTEx used both depth of coverage ratios and B allele frequencies calculated from whole exome sequencing data, to predict copy number variations along with their genotypes [88]. More and more new algorithms and databases have been established to provide the basis for the application of pattern recognition.

Prospective usage of pattern recognition for PPPM in cancer

The incidence of cancer increased year by year, and more and more people die of cancer. Because there is a huge difference in the 5-year survival rate of early treatment and late treatment, so early diagnosis is particularly necessary. We can use the gene pattern derived from high-risk group to perform risk assessment, and improve cancer screening, early diagnosis, and treatment.

Due to the tumor heterogeneity, different patients have different gene mutations, which lead to different sensitivity to the drug. Thus identity of differentially expressed genes was needed for precise treatment. Some effective cancer biomarkers have been discovered and used in clinic. For example, CEA and AFP are the most common tumor markers that are derived from abnormal protein products of tumor cells. However, due to low specificity of these proteins, it only plays a supporting role, but not a determing factor in clinic diagnosis. With further studies, more and more differentially expressed proteins or peptides will be found; these proteins or peptides combined to form a pattern, increase specificity of the tumor diagnosis, and reduce the false positive rate.

The pattern that mentioned above could be composed by different types of biomarkers from genome, transcriptome, proteome, metabolome, and radiome. Not only the same kind of molecular markers can be composed of pattern, different kinds of molecular markers can also be combined together to form an integrative pattern, for example, mass spectrometry imaging data and gene expression microarray data are composed into an integrative pattern. Analysis results show that a patten that combined MSI data and biological data is able to provide a meaningful discrimination between samples. It might be a useful tool to identify potential in large-scale biological, especially to identify cancer patient and health people [72].

However, there are still some problems regarding pattern recognition. First, it perhaps has different variations of genes or proteins in the different stages of tumor development. How to identify these genes and their proteins remains a challenge. Second, the recurrence of tumor is not only a simple change of gene or protein, but also is closely related to the patient's living environment and eating habits. Only focus on one aspect is not enough. In the future, one has to combine these laboratory parameters with the patients' daily habits together to create a pattern model, in order to achieve a more accurate prediction of tumor and individualized treatment. Combined with other factors, such as age, sex, family history, obesity, lifestyle, etc. The model one expects to establish is a series of data from patients which can predict the probability of occurrence of a tumor, and is able to change specific medications according to key sites. It is necessary to establish a model for prediction, prognosis and the best choice of drug use for cancer patients (Fig. 3).

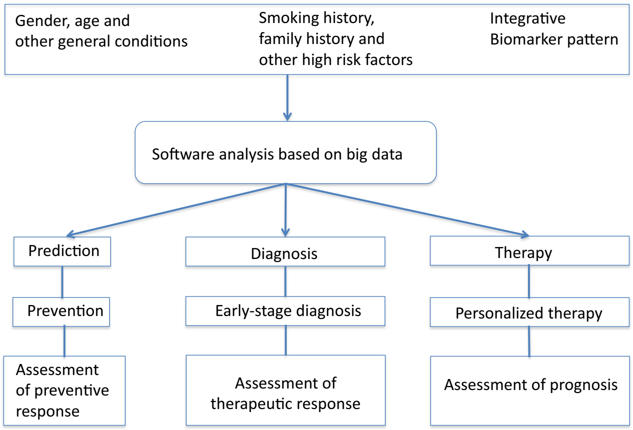
Gender, age and other general conditions
Smoking history, family history and other high risk factors
Integrative Biomarker pattern
Software analysis based on big data
Prediction
Diagnosis
Therapy
Prevention
Early-stage diagnosis
Personalized therapy
Assessment of preventive response
Assessment of therapeutic response
Assessment of prognosis

Fig. 3 Ideal model about pattern recognition

Conclusion

Precision medicine requires us to do early diagnosis and individualized treatment, and improve the specificity of diagnosis and treatment. The traditional single biomarker prediction model is very difficult to have higher sensitivity and specificity, so there is a need to form the biomarker pattern. The development in DNA, RNA, protein, and imaging techniques offers promise to find more biomarker pattern. Pattern recognition can not only be between the same kind of pattern, but also can be between different categories, such as some DNA biomarkers and cancer imaging features together to form a pattern. In addition, the progress of computer technology and the emergence of the new algorithm provide one the possibility to realize the pattern recognition. A pattern recognition model is expected to build and realize the early diagnosis, accurate prognostic evaluation, and selection of better drugs for cancer patients.

Acknowledgements This work was supported by the grants from China "863" Plan Project (Grant No. 2014AA020610-1 to XZ), the Xiangya Hospital Funds for Talent Introduction (to XZ), the National Natural Science Foundation of China (Grant No. 81272798 and 81572278 to XZ), and the Hunan Provincial Natural Science Foundation of China (Grant No. 14JJ7008 to XZ).

References

1. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz Jr LA, Kinzler KW. Cancer genome landscapes. Science. 2013;339:1546–58.
2. Hoth M. CRAC channels, calcium, and cancer in light of the driver and passenger concept. Biochim Biophys Acta. 2016;1863(6 Pt B): 1408–17.
3. Liu J, Mao Y, Li Z, Zhang D, Zhang Z, Hao S, et al. Use of texture analysis based on contrast-enhanced MRI to predict treatment response to chemoradiotherapy in nasopharyngeal carcinoma. J Magn Reson Imaging. 2016;44:445–55.
4. Derks S, Cleven AH, Melotte V, Smits KM, Brandes JC, Azad N, et al. Emerging evidence for CHFR as a cancer biomarker: from tumor biology to precision medicine. Cancer Metastasis Rev. 2014;33:161–71.
5. Liu FF. Novel gene therapy approach for nasopharyngeal carcinoma. Semin Cancer Biol. 2002;12:505–15.
6. Gregory TR. Synergy between sequence and size in large-scale genomics. Nat Rev Genet. 2005;6:699–708.
7. Jorde LB, Wooding SP. Genetic variation, classification and 'race'. Nat Genet. 2004;36(11 Suppl):S28–33.
8. Tan DS, Mok TS, Rebbeck TR. Cancer genomics: diversity and disparity across ethnicity and geography. J Clin Oncol. 2016;34:91–101.
9. Mroz EA, Rocco JW. The challenges of tumor genetic diversity. Cancer. 2016. doi:10.1002/cncr.30430.
10. Müller B, Wilcke A, Boulesteix AL, Brauer J, Passarge E, Boltze J, et al. Improved prediction of complex diseases by common genetic markers: state of the art and further perspectives. Hum Genet. 2016;135:259–72.
11. Cheon S. Probability concepts and distributions for analyzing large biological data. In: Lee JK, editor. Statistical bioinformatics for biomedical and life science researchers. Hoboken: Willey; 2010. p. 7–56.
12. Gray KA, Yates B, Seal RL, Wright MW, Bruford EA. Genenames.org: the HGNC resources in. Nucleic Acids Res 2015. 2015;43(Database issue):D1079–85.
13. Pique-Regi R, Monso-Varona J, Ortega A, Seeger RC, Triche TJ, Asgharzadeh S. Sparse representation and Bayesian detection of genome copy number alterations from microarray data. Bioinformatics. 2008;24:309–18.
14. Chiu CG, Nakamura Y, Chong KK, Huang SK, Kawas NP, Triche T, et al. Genome-wide characterization of circulating tumor cells identifies novel prognostic genomic alterations in systemic melanoma metastasis. Clin Chem. 2014;60:873–85.
15. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29:308–11.
16. Kweekel DM, Gelderblom H, Van der Straaten T, Antonini NF, Punt CJ, Guchelaar HJ. UGT1A1*28 genotype and irinotecan dosage in patients with metastatic colorectal cancer: a Dutch Colorectal Cancer Group study. Br J Cancer. 2008;99:275–82.
17. Gil J, Sąsiadek MM. Gilbert syndrome: the UGT1A1*28 promoter polymorphism as a biomarker of multifactorial diseases and drug metabolism. Biomark Med. 2012;6:223–30.
18. Zintzaras E, Ziogas DC, Kitsios GD, Papathanasiou AA, Lau J, Raman G. MTHFR gene polymorphisms and response to chemotherapy in colorectal cancer: a meta analysis. Pharmacogenomics. 2009;10:1285–94.
19. Xie Q, Ratnasinghe LD, Hong H, Perkins R, Tang ZZ, Hu N, et al. Decision forest analysis of 61 single nucleotide polymorphisms in a case-control study of esophageal cancer. BMC Bioinform. 2005;6 Suppl 2:S4.
20. Martinez VD, Vucic EA, Thu KL, Pikor LA, Hubaux R, Lam WL. Unique pattern of component gene disruption in the NRF2 inhibitor KEAP1/CUL3/RBX1 E3-ubiquitin ligase complex in serous ovarian cancer. BioMed Res Int. 2014;2014:159459.
21. Chen CP, Fushing H, Atwill R, Koehl P. biDCG: a new method for discovering global features of DNA microarray data via an iterative re-clustering procedure. PLoS One. 2014;9:e102445.
22. Sorenson GD, Pribish DM, Valone FH, Memoli VA, Bzik DJ, Yao SL. Soluble normal and mutated DNA sequences from single-copy genes in human blood. Cancer Epidemiol Biomarkers Prev. 1994;3:67–71.
23. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med. 2014, 6: 224ra
24. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. Nat Med. 2008;14:985–90.
25. Newman AM, Bratman SV, To J, Wynne JF, Eclov NC, Modlin LA, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med. 2014;20:548–54.
26. Cheng F, Su L, Qian C. Circulating tumor DNA: a promising biomarker in the liquid biopsy of cancer. Oncotarget. 2016;7:48832–41.
27. Alahari SV, Eastlack SC, Alahari SK. Role of long noncoding RNAs in neoplasia: special emphasis on prostate cancer. Int Rev Cell Mol Biol. 2016;324:229–54.
28. Wu W, Sun M, Zou GM, Chen J. MicroRNA and cancer: current status and prospective. Int J Cancer. 2007;120:953–60.
29. Wang Y, Stricker HM, Gou D, Liu L. MicroRNA: past and present. Front Biosci. 2007;12:2316–29.
30. Chen PY, Meister G. MicroRNA-guided posttranscriptional gene regulation. Biol Chem. 2005;386:1205–18.
31. Alvarez-Garcia I, Miska EA. MicroRNA functions in animal development and human disease. Development. 2005;132:4653–62.
32. Gregory RI, Shiekhattar R. MicroRNA biogenesis and cancer. Cancer Res. 2005;65:3509–12.
33. Lehmann U. Aberrant DNA methylation of microRNA genes in human breast cancer a critical appraisal. Cell Tissue Res. 2014;356:657–64.
34. Xiao YF, Yong X, Fan YH, Lü MH, Yang SM, Hu CJ. MicroRNA detection in feces, sputum, pleural effusion and urine: novel tools for cancer screening (review). Oncol Rep. 2013;30:535–44.
35. de Planell-Saguer M, Rodicio MC. Analytical aspects of microRNA in diagnostics: a review. Anal Chimica Acta. 2011;699:134–52.
36. Andorfer CA, Necela BM, Thompson EA, Perez EA. MicroRNA signatures: clinical biomarkers for the diagnosis and treatment of breast cancer. Trends Mol Med. 2011;17:313–9.
37. Zoon CK, Starker EQ, Wilson AM, Emmert-Buck MR, Libutti SK, Tangrea MA. Current molecular diagnostics of breast cancer and the potential incorporation of microRNA. Expert Rev Mol Diagn. 2009;9:455–67.
38. Deng S, Calin GA, Croce CM, Coukos G, Zhang L. Mechanisms of microRNA deregulation in human cancer. Cell Cycle. 2008;7: 2643–6.
39. Zhang L, Yang N, Coukos G. MicroRNA in human cancer: one step forward in diagnosis and treatment. Adv Exp Med Biol. 2008;622: 69–78.
40. Kosaka N, Iguchi H, Ochiya T. Circulating microRNA in body fluid: a new potential biomarker for cancer diagnosis and prognosis. Cancer Sci. 2010;101:2087–92.
41. Wang J, Zhang KY, Liu SM, Sen S. Tumor-associated circulating microRNAs as biomarkers of cancer. Molecules. 2014;19:1912–38.
42. Farina NH, Wood ME, Perrapato SD, Francklyn CS, Stein GS, Stein JL, et al. Standardizing analysis of circulating microRNA: clinical and biological relevance. J Cel Biochem. 2014;115:805–11.
43. Schwarzenbach H. Circulating nucleic acids as biomarkers in breast cancer. Breast Cancer Res. 2013;15:211.
44. Ma R, Jiang T, Kang X. Circulating microRNAs in cancer: origin, function and application. J Exp Clin Cancer Res. 2012;31:38.
45. Yu DC, Li QG, Ding XW, Ding YT. Circulating microRNAs: potential biomarkers for cancer. Int J Mol Sci. 2011;12:2055–63.

46. Mostert B, Sieuwerts AM, Martens JW, Sleijfer S. Diagnostic applications of cell-free and circulating tumor cell-associated miRNAs in cancer patients. Expert Rev Mol Diagn. 2011;11:259–75.

47. Heneghan HM, Miller N, Kelly R, Newell J, Kerin MJ. Systemic miRNA-195 differentiates breast cancer from other malignancies and is a potential biomarker for detecting noninvasive and early stage disease. Oncologist. 2010;15:673–82.

48. Vlassov VV, Laktionov PP, Rykova EY. Circulating nucleic acids as a potential source for cancer biomarkers. Cur Mol Med. 2010;10:142–65.

49. Heneghan HM, Miller N, Lowery AJ, Sweeney KJ, Newell J, Kerin MJ. Circulating microRNAs as novel minimally invasive biomarkers for breast cancer. Ann Surg. 2010;251:499–505.

50. Zhu W, Qin W, Atasoy U, Sauter ER. Circulating microRNAs in breast cancer and healthy subjects. BMC Res Notes. 2009;2:89.

51. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci U S A. 2008;105:10513–8.

52. Zhao H, Shen J, Medico L, Wang D, Ambrosone CB, Liu S. A pilot study of circulating miRNAs as potential biomarkers of early stage breast cancer. PLoS One. 2010;5:e13735.

53. Noren Hooten N, Fitzpatrick M, Wood 3rd WH, De S, Ejiogu N, Zhang Y, et al. Age-related changes in microRNA levels in serum. Aging (Albany NY). 2013;5:725–40.

54. Olivieri F, Rippo MR, Procopio AD, Fazioli F. Circulating inflamma-miRs in aging and age-related diseases. Front Genet. 2013;4:121.

55. Mishra PJ. MicroRNA polymorphisms: a giant leap towards personalized medicine. Per Med. 2009;6:119–25.

56. Wu X, Weng L, Li X, Guo C, Pal SK, Jin JM, et al. Identification of a 4-microRNA signature for clear cell renal cell carcinoma metastasis and prognosis. PLoS One. 2012;7:e35661.

57. Mishra PJ. Non-coding RNAs as clinical biomarkers for cancer diagnosis and prognosis. Expert Rev Mol Diagn. 2014;14:917–9.

58. Mishra PJ. MicroRNAs as promising biomarkers in cancer diagnostics. Biomark Res. 2014;2:19.

59. Houseley J, Rubbi L, Grunstein M, Tollervey D, Vogelauer M. A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. Mol Cell. 2008;32:685–95.

60. Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, Levin JZ, et al. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. Genomes Res. 2012;22:577–91.

61. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. Cell. 2009;136:629–41.

62. Hao Y, Crenshaw T, Moulton T, Newcomb E, Tycko B. Tumour-suppressor activity of H19 RNA. Nature. 1993;365:764–7.

63. Shen J, Hu Q, Schrauder M, Yan L, Wang D, Medico L, et al. Circulating miR-148b and miR-133a as biomarkers for breast cancer detection. Oncotarget. 2014;5:5284–94.

64. Jungblut PR, Holzhütter HG, Apweiler R, Schlüter H. The speciation of the proteome. Chem Cent J. 2008;2:16.

65. Zhan X, Giorgianni F, Desiderio DM. Proteomics analysis of growth hormone isoform in the human pituitary. Proteomics. 2005;5:1228–41.

66. Black DL. Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem. 2003;72:291–336.

67. Kohler M, Thomas A, Pushel K, Schanzern W, Thevis M. Identification of human pituitary growth hormone variants by mass spectrometry. J Proteome Res. 2008;8:1071–6.

68. Okkels LM, Müller EC, Schmid M, Rosenkrands I, Kaufmann SH, Andersen P, et al. CFP10 discriminates between nonacetylated and acetylated ESAT-6 of Mycobacterium tuberculosis by differential interaction. Proteomics. 2004;4:2954–60.

69. Schlüter H, Apweiler R, Holzhütter HG, Jungblut PR. Finding one's way in proteomics: a protein species nomenclature. Chem Cent J. 2009;3:11.

70. Doetsch F. The glial identity of neural stem cells. Nat Neurosci. 2003;11:1127–34.

71. Li Z, Lu H, Yang J, Zeng X, Zhao L, Li H, et al. Analysis of the raw serum peptidomic pattern in glioma patients. Clin Chim Acta. 2013;425:221–6.

72. Kaddi CD, Parry RM, Wang MD. Multivariate hypergeometric similarity measure. IEEE/ACM Trans Comput Biol Bioinform. 2013;10:1505–16.

73. Veselkov KA, Mirnezami R, Strittmatter N, Kinross J, Speller A, Abramov T, et al. Chemo-informatic strategy for imaging mass spectrometry-based hyperspectral profiling of lipid signatures in colorectal cancer. Proc Natl Acad Sci U S A. 2014;111:1216–21.

74. Lee JH, Kim KH, Park JW, Chang HJ, Kim BC, Kim SY, et al. Low-mass-ion discriminant equation: a new concept for colorectal cancer screening. Int J Cancer. 2014;134:1844–53.

75. Lin XL, Yang SY, Du J, Tian YX, Bu LN, Huo SF, et al. Detection of lung adenocarcinoma using magnetic beads based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry serum protein profiling. Chin Med J (Engl). 2010;123:34–9.

76. Xu ZQ, Broza YY, Ionsecu R, Tisch U, Ding L, Liu H, et al. A nanomaterial-based breath test for distinguishing gastric cancer from benign gastric conditions. Br J Cancer. 2013;108:941–50.

77. Van Berkel JJ, Dallinga JW, Möller GM, Godschalk RW, Moonen E, Wouters EF, et al.. Development of accurate classification method based on the analysis of volatile organic compounds from human exhaled air. J Chromatogr B Analyt Technol Biomed Life Sci. 861: 101–107.

78. van de Kant KD, van der Sande LJ, Jöbsis Q, van Schayck OC, Dompeling E. Clinical use of exhaled volatile organic compounds in pulmonary diseases: a systematic review. Respir Res. 2012;13:117.

79. Horváth I, Lázár Z, Gyulai N, Kollai M, Losonczy G. Exhaled biomarkers in lung cancer. Eur Respir J. 2009;34:261–75.

80. Hazelbag S, Fleuren GJ, Baelde JJ, Schuuring E, Kenter GG, Gorter A. Cytokine profile of cervical cancer cells. Gynecol Oncol. 2001;83:235–43.

81. Yamamoto T, Kimura T, Ueta E, Tatemoto Y, Osaki T. Characteristic cytokine generation patterns in cancer cells and infiltrating lymphocytes in oral squamous cell carcinomas and the influence of chemoradiation combined with immunotherapy on these patterns. Oncology. 2003;64:407–15.

82. Gardner JW, Bartlett PN. Applications and advances in electronic-nose technologies. Sensors (Basel). 2009;9:5099–148.

83. Mazzone PJ, Wang XF, Xu Y, Mekhail T, Beukemann MC, Na J, et al. Exhaled breath analysis with a colorimetric sensor array for the identification and characterization of lung cancer. J Thorac Oncol. 2012;7:137–42.

84. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. Magn Reson Imaging. 2012;30:1234–48.

85. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48:441–6.

86. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.

87. Zanin M, Menasalvas E, Boccaletti S, Sousa P. Feature selection in the reconstruction of complex network representations of spectral data. PLoS One. 2013;8:e72045.

88. Amarasinghe KC, Li J, Hunter SM, Ryland GL, Cowin PA, Campbell IG, et al. Inferring copy number and genotype in tumour exome data. BMC Genomics. 2014;15:732.