Seminars in Cancer Biology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Seminars in Cancer Biology

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

RNA sequencing for research and diagnostics in clinical oncology

Anton Buzdin[a,b,c,*], Maxim Sorokin[a,b,c], Andrew Garazha[b], Alexander Glusker[b], Alex Aleshin[d], Elena Poddubskaya[a,e], Marina Sekacheva[a], Ella Kim[f], Nurshat Gaifullin[g], Alf Giese[h], Alexander Seryakov[i], Pavel Rumiantsev[j], Sergey Moshkovskii[k,l], Alexey Moiseev[a]

[a] I.M. Sechenov First Moscow State Medical University, Moscow, Russia
[b] Omicsway Corp., Walnut, CA, USA
[c] Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia
[d] Stanford University School of Medicine, Stanford, 94305, CA, USA
[e] Vitamed Oncological Clinics, Moscow, Russia
[f] Johannes Gutenberg University Mainz, Mainz, Germany
[g] Lomonosov Moscow State University, Faculty of Medicine, Moscow, Russia
[h] Orthocenter Hamburg, Germany
[i] Medical Holding SM-Clinic, Moscow, Russia
[j] Endocrinology Research Center, Moscow, 117312, Russia
[k] Institute of Biomedical Chemistry, Moscow, 119121, Russia
[l] Pirogov Russian National Research Medical University (RNRMU), Moscow, 117997, Russia

ARTICLE INFO

Keywords:
Transcriptomics
RNA sequencing
Genomics
DNA mutations
Molecular diagnostics
Recurrent and metastatic disease
Clinical oncology
Targeted therapies
Personalized medicine
Molecular markers
Bioinformatics

ABSTRACT

Molecular diagnostics is becoming one of the major drivers of personalized oncology. With hundreds of different approved anticancer drugs and regimens of their administration, selecting the proper treatment for a patient is at least nontrivial task. This is especially sound for the cases of recurrent and metastatic cancers where the standard lines of therapy failed. Recent trials demonstrated that mutation assays have a strong limitation in personalized selection of therapeutics, consequently, most of the drugs cannot be ranked and only a small percentage of patients can benefit from the screening. Other approaches are, therefore, needed to address a problem of finding proper targeted therapies. The analysis of RNA expression (transcriptomic) profiles presents a reasonable solution because transcriptomics stands a few steps closer to tumor phenotype than the genome analysis. Several recent studies pioneered using transcriptomics for practical oncology and showed truly encouraging clinical results. The possibility of directly measuring of expression levels of molecular drugs' targets and profiling activation of the relevant molecular pathways enables personalized prioritizing for all types of molecular-targeted therapies. RNA sequencing is the most robust tool for the high throughput quantitative transcriptomics. Its use, potentials, and limitations for the clinical oncology will be reviewed here along with the technical aspects such as optimal types of biosamples, RNA sequencing profile normalization, quality controls and several levels of data analysis.

1. Introduction: a brief overview of the targeted therapy in cancer

The turn of Millennium was marked by the series of spectacular breakthroughs in molecular medicine. The completion of Human Genome Project coincided with the development of three game-changing therapies in oncology, namely, imatinib (inhibitor of fusion BCR-ABL tyrosine kinase) in chronic myelogeneous leukemia (the same drug was later found to be active as a KIT inhibitor in gastrointestinal stromal tumors), rituximab (monoclonal antibodies against CD20) in B-cell lymphoma, and trastuzumab (anti-HER2 antibodies) in breast cancer [1–3]. A successful treatment of these cancers, refractory to traditional cytostatic chemotherapy, accompanied by a moderate toxicity, was welcomed by the term "targeted therapy", contrasting the new drugs, aiming at molecules specific (exclusively or mainly) to the tumor cells, with "old", molecularly indiscriminative and, consequently, toxic chemotherapy, inflicting severe damage to many healthy tissues. Indeed, the prospects emerged to find corresponding targets in all or most tumors and, eventually, to score a victory in the fight against cancer.

Medical and pharmaceutical community rushed into the proverbial breach, and further molecularly targeted drugs were developed and

* Corresponding author at: 340 S Lemon Ave, 6040, Walnut, 91789 CA, USA.
E-mail address: buzdin@oncobox.com (A. Buzdin).

https://doi.org/10.1016/j.semcancer.2019.07.010
Received 30 June 2019; Accepted 16 July 2019

entered clinical trials, among others including inhibitors of RAS and RAF proteins, antiapoptotic BCL2, vascular endothelial growth factor (VEGF), and its receptors, matrix metalloproteinases (MMPs), epidermal growth factor receptor (EGFR) [4–6]. Several agents such as sorafenib and sunitinib were designed to aim at multiple targets and hence called “multitargeted drugs”. However, the success of these approaches was limited at best, and many target molecules deemed non-druggable. Tumor responses were rare and mostly short-term, and toxicity, while different from those of chemotherapy, became significant. Thus, the distinction between targeted drugs and chemotherapy became blurred, and the optimism toward curing cancer waned.

One of the main targets for anticancer treatment became tumor angiogenesis. Monoclonal antibodies to VEGF (bevacizumab), VEGF-trap (aflibercept) and antibodies to VEGF receptor 2 (ramucirumab) as well as numerous small-molecule inhibitors of VEGF receptors were extensively studied in practically all tumor types. The results were mixed; even in renal cancer, clearly dependent on angiogenesis, durable responses were infrequent. The best strategy that unexpectedly (and somehow counter intuitively) emerged from these trials was to combine bevacizumab with cytotoxic chemotherapy, only to improve overall survival by 10–15% (1.5–3 months in most cases) [7,8].

After several years of trial and error, it became clear that the target must be defined more meticulously. For example, EGFR inhibitors gefitinib and erlotinib failed to improve survival in non-selected lung cancer populations, but were found to be active against rare tumors (mainly adenocarcinomas) with activating mutations in *EGFR* gene, namely L858R and exon 19 deletions [9–11]. Monoclonal antibodies to EGFR (cetuximab and panitumumab) showed some activity in colorectal cancer, but only in tumors without *KRAS*, *NRAS*, and *BRAF* mutations [12,13]. Many other tumors, like lung, esophageal, or head and neck cancers, have very little sensitivity to these agents, despite strong expression of EGFR. Different tumors with activated *BRAF* oncogene bearing V600E mutation were resistant to sorafenib, but amenable to the treatment with more selective drugs such as vemurafenib and dabrafenib. In fact, their efficacy was very tumor-specific, with strong and lasting responses in melanoma (particularly in combination with MEK inhibitors like trametinib), a decent effect in some other *BRAF*-mutated cancers (e.g., small fractions of glioblastomas, lung adenocarcinomas, and cholangiocarcinomas), but a complete lack of benefit in colorectal and thyroid cancers, there *BRAF* mutations are abundant [14–18].

Several relatively rare genetic abnormalities were described as targets for anticancer therapy, the most prominent examples in solid tumors being *ALK* and *ROS1* gene rearrangements in lung adenocarcinoma, susceptible to inhibition with crizotinib and its analogues [19,20], and sonic hedgehog pathway alterations in basal-cell carcinoma and medulloblastoma, successfully targeted with vismodegib [21–23]. Some important targets were detected at low frequencies in many different tumors, such as *NTRK* genes fusions in a broad range of cancers (including thyroid, colorectal, and lung adenocarcinomas, and even rare sarcomas and glioblastomas), inhibited by entrectinib and loratrectinib, as well as *MET* exon 14 skipping in several tumors, making them vulnerable to MET inhibitors like crizotinib or cabozantinib [24,25].

Further incremental advances in targeted therapy stemmed from the study of hereditary breast and ovarian cancers. Germ-line and somatic defects in *BRCA1*, *BRCA2*, *CHEK2*, *ATM* and other DNA damage repair genes found in these tumors (and at smaller percentages in several others, e.g. prostate and pancreatic cancers) make them amenable to synthetic lethality by using poly(ADP-ribose) polymerase (PARP) inhibitors, e.g. olaparib or veliparib [26].

Unfortunately, as advances in DNA sequencing defined molecular portraits of most cancer subtypes and comprehensive genomic profiles of common tumors became available, the proportion of cancer patients with actionable molecular targets, where currently available drugs can achieve a durable remission, appeared to shrink to 10–15%.

A landmark achievement that practically doubled that dismal numbers was the development of immune checkpoint inhibitors, namely, antibodies cytotoxic T-lymphocyte–associated antigen 4 (CTL4), programmed cell death protein 1 (PD-1), and PD-L1. These agents — antibodies to CTLA4 (ipilimumab), PD-1 (nivolumab and pemprolizumab) and PD-L1 (atezolizumab, durvalumab and avelumab) — may induce strong responses in many tumor types, but, in contrast to several targeted drugs described above, robust predictors of their efficacy are still elusive, with a single exception being microsatellite instability in tiny fraction of different tumors, mainly colorectal cancer [27–30].

Overall, despite spectacular successes in a significant, but still small percentage of patients, targeted therapy in oncology faces several challenges: a lack of genomic (mutation-based) predictors of efficacy for many agents, development of resistance after the initial response in sensitive tumors, as well as tumor heterogeneity with a limited ability to predict the fate of metastases after single primary biopsy [29,31].

Personalized diagnostics is viewed as a future standard in oncology. Currently, however, only a limited number of genetic platforms use several types of high-throughput genetic profiling to consult physicians and patients about targeted therapy. The examples of such platforms are Caris Molecular Intelligence and Foundation One [32–34]. Their use to obtain clinically actionable results is based on an analysis of a defined spectrum of mutations with previously demonstrated significance, as well as on immunohistochemical (IHC) profiling of several protein markers. Resulting mutational profiles are limited by a panel of target genes (overwhelming minority of the total number) and are used mainly to calculate overall tumor mutational burden. By this approach, most of obtained genetic information is of no use. Cited platforms do not use high-throughput expression profiles, do not analyze upregulation of signaling pathways and are not suited for integration of multiomics data. Consequently, their analytical power is considerably restricted.

This article examines the approaches that apply transcriptomic analysis to overcome these obstacles and improve targeted therapy selection in solid tumors. We assume that significant progress in this area may be attainable by exploring up- and downregulation of the intracellular signaling pathways using integrative data from the deep exomic and transcriptomic sequencing.

2. Profiling of gene expression in clinical oncology

Gene expression can be monitored on both mRNA and protein levels [35] while both dimensions have their advantages and limitations. The protein level is obviously closer to the cellular or tissue phenotype because these are the proteins that execute major molecular functions in a living cell (Fig. 1). There is a plethora of methods for measuring expression levels for single proteins in fresh or fixed tumor cells and tissues, including immunoassays such as immunohistochemistry, Western blotting and other super hits in modern clinical lab diagnostics [36]. For better performance of cancer diagnostic and predictive classifiers, immunoassays are multiplexed [37]. At the same time, the number of techniques for high throughput and analytically robust protein quantitation is quite limited. Primarily, this can be performed by means of protein microarrays [38] and mass spectrometry-linked proteomic assays [39]. In hunt for new diagnostic and theranostic methods, hundreds and thousands of proteins are intended to be measured with microarrays or mass-spectrometry at the discovery phase. Then, if the more limited diagnostic signature is selected, multiplex immunoassays or targeted mass-spectrometric tests may be further used in clinical trials [40].

3. Protein microarrays

Protein microarrays are based on the principle of specific

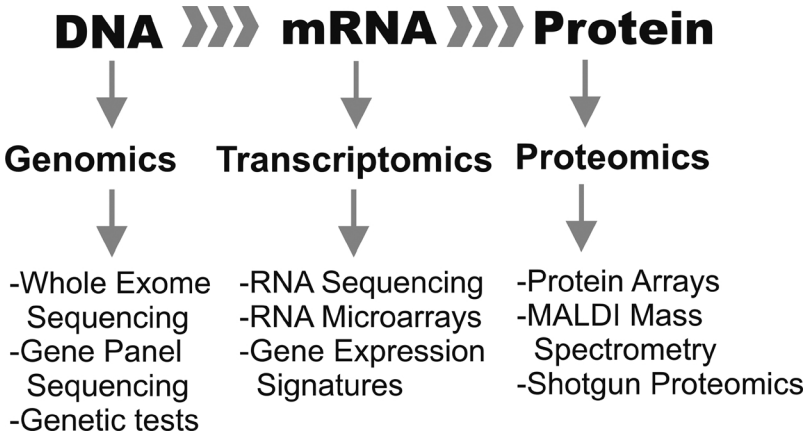

Fig. 1. Outline of major OMICS approaches in clinical oncology.

recognition of proteins by the respective monoclonal antibodies or other binders, such as oligonucleotide aptamers, where each sector on an array is reserved to an individual protein. The principle of antibody microarray fabrication is known from early 1980s [41], and since that time, they were often used for cancer biomarker discovery [42,43]. Besides the arrays of antibodies linked to the solid phase, reverse phase microarrays are widely used, where small tissue samples are attached to the surface which is then probed by individual antibody in each spot [44]. The antibody microarray approach has several technical difficulties. First of all, there is a need in highly specific and sensitive panel of monoclonal antibodies to interrogate complete human proteome or its significant part. This is technically challenging, especially because conformation/modification patterns of individual proteins from human tissues may differ dramatically from those used *in vitro* to obtain the analytic antibodies [38]. The challenge of antibody handling is partly met by use of microarrays, where native or modified aptamers are used as multiple protein binders [45]. For example, a technical platform often used in cancer research is a proprietary Somascan™ modified aptamer-based microarray [46]. Use of the latter started from blood plasma applications [47] and now proceeds to cancer tissue assays [48].

Another major problem for any protein microarrays is different physical and chemical properties like solubility and hydrophobicity of different proteins extracted from the tissues to be analyzed [49]. Finally, different proteins are degraded with different rates, especially in different tissues, and all the technical procedures like sample isolation, protein extraction and processing must be strictly synchronized to attempt obtaining meaningful data [49]. Taken together this creates substantial obstacles to work with minute amounts of clinical biomaterials, to reproducibly measure and compare protein levels [50].

4. Mass-spectrometry – initial MALDI profiling

Besides protein microarray, the second major branch of methods for multiple protein identification and quantitation of cancer-related samples is a mass-spectrometry-based proteomics [51]. In the very beginning of proteomics itself, in early 2000s, a first approach to distinguish cancerous specimens based on their protein composition was blood plasma profiling using matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) mass spectrometry [52]. In this approach, blood plasma or other biosample is applied to the metal surface almost without separation, and tens to hundreds of proteins are detected. The resultant profiles may be compared between normal and cancer samples and were thought to serve as diagnostic assays for clinical applications [53]. It was shown soon after introduction of the approach that these profiles lack analytical sensitivity and thus formed by highly abundant peptides and proteins, mostly the acute phase reactants [54]. Thus, without enough diagnostic accuracy, MALDI-TOF proteomic profiles were not further considered for primary cancer diagnostics or screening. However, multiple efforts were made to translate the MALDI-TOF based assays into clinical practice as prognostic and/or predictive tests [55]. The most successful of them is a currently marketed Veristrat test which is prognostic for outcome of non-small cell lung cancer (NSCLC) with the erlotinib treatment and predictive of differential treatment benefit between erlotinib and chemotherapy [56].

5. Shotgun proteomics

After development of MALDI-TOF-based assays, which met criticism for their insufficient sensitivity, the shotgun proteomics was introduced for analysis of cancer cells and tissues. In the shotgun approach, the proteins extracted from a biosample are completely digested by proteolytic enzyme(s) with known specificity(ies) thus generating a characteristic repertoire of peptides that can be resolved as specific peaks following high-resolution mass spectrometry [39]. These may serve to characterize protein contents and abundances in biosamples. The state-of-the-art shotgun proteomic workflow may identify and even quantify up to ten thousand gene products from one sample [57]. Many examples can be mentioned of these techniques' applicability to discovery and tracing of cancer biomarkers [58–60], in characterizing functional dynamics of cancer proteome, such as changes in protein phosphorylation or other types of modifications [61,62].

6. FFPE versus frozen tissue

There is an attractive option to use formalin-fixed, paraffin embedded (FFPE) tumor specimens instead of freshly frozen tissues in biomarker discovery phase of mass-spectrometric assay. The proportion of papers which describe such a use of FFPE tissue is growing [63]. In the field, a discussion has been started if the analysis of archived tissues may serve an appropriate alternative to frozen samples [64]. A decade ago, a major study demonstrated that these two types of cancer samples can be used for shotgun proteomic assays interchangeably [65]. However, it is known that the procedure of embedding and consequent release of proteins from the paraffin leads to irreversible covalent changes of native protein sequences [64]. With a progress in sensitivity of protein identification, these changes may influence to the retrieval of low-abundant proteins which are of potential importance for cancer [63]. Thus, the major CPTAC cancer proteomic consortium uses frozen microdissected tissues for the proteogenomic projects which add proteomic data to the cancer genomes and transcriptomes [66,67]. At the same time, FFPE tumor samples are widely used in proteomic studies with shotgun [68] and targeted [69] mass-spectrometric methods, as well as for MALDI-based mass-spectrometric tissue imaging [70].

7. Proteogenomic approach

In molecular cancer studies, it is attractive to combine results of two or more omics techniques for exhaustive characterization of each tumor sample. The approach of combining deep nucleic acid sequencing and shotgun proteomics was generally called proteogenomics. The latter made it possible to better classify cases of the same tumor type for prognostic and predictive decisions [71] and also to identify mutant proteins as biomarkers and neoantigens [72].

Major proteogenomic studies which included hundreds of individual tumors were started by CPTAC consortium for the major cancer types, such as colon, breast and ovarian cancer [66,73,74]. The data integration provided clinically relevant results. For example, the recent proteogenomic study of colon cancer allowed identification of potential protein signatures predictive for response to the immune checkpoint inhibitor therapy [67]. Proteomics therefore is on its way to becoming a method of choice for profiling gene expression in clinical applications, although several issues like reproducibility, costs and (not) ease of use are still to be solved.

8. Ribosomal profiling

On the interface between proteomics and transcriptomics, another approach termed “translatomics” [75], or *ribosomal profiling* should be mentioned based on isolation of mRNA molecules directly bound with ribosome followed by their deep sequencing [76]. This makes it possible to quantitatively estimate translation of mRNAs by ribosomes in a biosample [77]. This approach is one step closer to the cancer phenotype than transcriptomics itself, thus attracting growing attention in biomedical research [78]. However, the procedure of ribosome profiling is still laborious and experimentally challenging, thus limiting its broad applicability [79]. It is also impossible for fixed clinical biomaterials, because it requires substantial amounts of fresh tissue which is problematic for clinical diagnostics of most of solid cancers [80].

9. Transcriptomics

Transcriptomics deals with directly analyzing concentrations of RNA molecules in biosamples, including mRNAs for protein coding genes and microRNAs [81]. Standing at a longer distance from the phenotype than translatomics, transcriptomics remains unparalleled approach in terms of generating high throughput gene expression data using robust and relatively chip experimental protocols suitable for the analysis of fresh and fixed clinical biomaterials [82]. A growing transcriptomic data is deposited in the special public repositories like Gene Expression Omnibus, GEO [83] and Array-Express [84]. accumulating more than two million of individual profiles obtained in more than 100,000 series of experiments [85].

Importantly, strong statistically significant correlations (r range 0.59-0.89) have been reported between ribosomal profiling and transcriptomic data for mRNA molecules [86–88], thus justifying the use of transcriptomics for functional interrogation of gene expression.

Three major directions in cancer transcriptomics can be mentioned such as quantitative polymerase chain reaction (PCR) [89], microarray hybridization [90] and RNA sequencing [91]. In quantitative PCR assays, simultaneous profiling of multiple gene expressions can be made at a very high accuracy [92], but unfortunately this type of gene screening cannot be done on a truly high-performance scale because of technical limitations. PCR-on-chip systems are the most labor saving among this group of methods, but they are limited by simultaneously probing hundreds or up to one thousand of human genes, e.g. [93].

Microarray hybridization is based on complementary hybridization of nucleic acids in solution, where the extent of gene product bound to a surface-immobilized specific probe is proportionate to its concentration in a biosample [94]. Multiple microarray platforms have been used for analyzing gene expression in human cancers, including Affymetrix [95], Illumina [96], Agilent [97] and CustomArray [98] systems. They utilize different library preparation protocols, e.g. different enzymes and numbers of PCR cycles, but also use different equipment and physical principles for detecting hybridization signals [99]. The same is true also for the different RNA sequencing platforms, such as Illumina [100], Ion Torrent/Proton [100] and Oxford Nanopore [101] systems.

These differences in sample preparation and screening methods result in a dramatic incompatibility of the results obtained using different platforms, reagents and kits, both for the microarray family and RNA sequencing techniques [102,103]. This is the reason why experimental results on gene expression are generally compared within the same platform [103].

Performances of RNA sequencing and expression microarrays were compared in many published investigations [104–106]. This is generally accepted now that RNA sequencing demonstrates superior precision in measuring gene expression than microarray hybridization, e.g. lower false discovery rate for differentially expressed genes [93,107]. Historically, microarray technology was introduced to the field in mid-nineties [108] and RNA sequencing – only about a dozen years later [109]. This resulted in strong penetration of microarray analyzes in various research protocols, so that major portion of currently publicly available gene expression datasets were obtained using various microarray platforms [110]. However, nowadays RNA sequencing is considered a gold standard for high throughput screening of gene expression, for all types of clinical biomaterial [107].

10. Data normalization and incompatibility of RNA sequencing information

As mentioned in the previous section, technical aspects like protocols and sequencing platforms used result in a compromised compatibility of RNA sequencing data. This has a functional consequence that the data to be directly compared ideally must be obtained using the same equipment, protocols and reagents. In many clinical oncology applications, gene expression is compared in tumor samples with the normal samples, e.g. [111–114]. Attention, therefore, should be paid to investigate compatibility of the data under comparison. For example, several collections (atlases) of RNA sequencing profiles for normal human tissues have been published. Ideally, they must represent normal tissues from healthy donors, profiled in a single series of experiments using the same equipment and reagents. The largest published dataset, GTEx [115] (11,688 samples), lacks publicly available data on the donor’s age, which limits its use for many practical applications in cancer biology. Access to the primary GTEx data also requires sophisticated registration steps that cannot be performed by every investigator. The other relevant databases of a significant size include the information on the donor’s age: TCGA [116] (625 samples), ENCODE [117] polyA RNA-seq (41 samples) and ENCODE total RNA-seq (92 samples). However, they lack one or several of the previously mentioned features. For example, in The Cancer Genome Atlas (TCGA) project database, the norms are considered specimens of histologically normal tissue adjacent to surgically removed tumors [118]. However, these tissues may be considered not physiologically normal due to numerous effects tumors may exert on the neighboring cells including biased growth factors and cytokine balances [119], pathological inflammation [120] and altered vascularization [121]. The ENCODE datasets were generated for the autopsy normal tissues by RNA sequencing using different library preparation methods. They have only 1–4 samples profiled per tissue type (both male and female donors included) and in most of the tissue types they can’t form a statistically significant reference group.

We recently constructed and deposited an alternative collection of original normal human tissue expression profiles obtained using Illumina HiSeq-3000 engine [122]. To this end, a total of 142 solid tissue samples representing 20 organs were taken from human healthy donors killed in road accidents. In addition, blood samples were

obtained from 17 healthy volunteers. Gene expression was profiled in one series of experiments using the same reagents and protocols (data ID GSE120795). When compared for compatibility with the transcriptomic data from the previous databases, we found that totally 399 profiles showed tissue-specific rather than platform- or database-specific clustering. All these data were collected in a database termed *Oncobox Atlas of Normal Tissue Expression (ANTE)* [122] including 11 sex-matched statistically significant tissue groups. The databases reviewed here may be useful to form reference groups during gene expression analyses in clinical biosamples.

The problem of gene expression data incompatibility makes problematic or even impossible comparison of experimental clinical profiles obtained using different experimental platforms and reagent sets [123–126]. For many published expression datasets there are available associated clinical outcomes linked with the patient's individual responses on cancer therapies [127]. Consequently, this non-comparability hampers further levels of data analysis for the different datasets, e.g. finding differentially expressed genes and assessing activation of molecular pathways [99,128].

Solving this problem would create a rich spectrum of clinically annotated molecular data, including profiles obtained for primary and metastatic cancer biosamples [103]. Accumulating these data in a comparable form would create an immense knowledge base that could be used for a plethora of practical applications in oncology such as drug discovery, biomarker development and formulation of combination therapies.

Accomplishing this task would require either initial gene expression data technical *normalization* (when datasets under comparison were obtained using the same experimental platform) or *harmonization* (when different platforms were used) [99,103]. The major normalization methods include quantile normalization (QN) [129], frozen robust multi-array analysis for microarray hybridization data (FRMA) [130], or DESeq [131] /DESeq2 [132]. There is also many methods reported suitable for harmonization including cross-platform normalization (XPN) [133,134], distance-weighted discrimination (DWD) [135,136], Empirical Bayes (EB) method also known as ComBat [137], Quantile Discretization (QD) [138], Normalized Discretization (NorDi) [138], DisTran (Distribution Transformation) [139], Gene Quantiles (GQ) [140], and platform-independent latent Dirichlet allocation (PLIDA) [141]. XPN method showed the best performance in a fundamental comparison of these harmonization techniques [133]. It acts by deeply restructuring distributions of gene expression levels for the samples under comparison. The algorithm uses data clustering to identify similarities between the gene expression profiles obtained using different platforms, and then expands these similarity regions by reshaping of the expression profiles. However, all harmonization methods mentioned above have a major limitation that they aren't capable of performing harmonization for more than two expression datasets [133]. They also show acceptable performance only for the datasets of a comparable sample size, therefore complicating harmonization of the real-world data [103]. Moreover, the resulting hybrid data are not further compatible with any of the existing formats for the same experimental platforms. This is a fundamental problem that doesn't allow converting various expression datasets into a uniform shape enabling multiple direct comparisons.

An attempt to solve this problem has been recently published [103]. A new cross-platform data harmonization technique termed *Shambhala* has been reported that is independent on (i) number of harmonized datasets and/or experimental platforms, and (ii) number of samples in every dataset. Shambhala harmonization converts in a universal predefined format the expression profiles taken one by one from the datasets under investigation, thus making them appropriate for further direct comparisons [103]. The initial version of this technique has been optimized for comparison of human transcriptomes, but it could be further developed to create a more universal tool useful for data analysis from many species. So far this approach has been successfully tested only for several model datasets [107] representing major microarray and RNA sequencing platforms [103]. Nevertheless, its further development may result in a transcriptomic analogue of BLAST search tool [142] that would be capable of ranking different transcriptomes by similarities using a plethora of available gene expression datasets. Such tool would clearly have strong practical impact in oncology as it might be used to identify "twin" transcriptomes, thus improving indications for therapy prescription and criteria for patient inclusion in clinical trials.

11. Multiple levels of RNA sequencing data analysis in oncology

Clinical RNA sequencing data can be analyzed in multiple ways, thus generating different outputs. Depending on the source of biomaterials, very different RNA sequencing results can be obtained. For appropriately stored fresh tissue specimens, high-integrity RNAs are isolated, thus leading to longer sequencing reads. For FFPE samples, significantly more degraded RNA preps can be obtained, typically resulting in 25–50 bp single end reads [122]. RNA sequencing reads are mapped on human genes and ambiguously mapping reads are removed from further analysis. Relative expression characteristics are then calculated for every gene. It should be noted that degraded RNAs can produce good-quality expression profiles that can form common clusters with the samples from high-integrity RNAs of the same biological origin [122]. Taking clustering of biologically similar samples as a measure of quality, a threshold of approx. 2.5 million RNA sequencing reads mapped on human protein coding genes was empirically established for the analysis of both fresh and FFPE human tissues [122]. Samples with lower number of reads didn't cluster in a biologically meaningful manner, and vice versa [122]. However, working with highly degraded RNAs from FFPE samples has a drawback that no analysis of fused oncogenes can be performed because of too short reads and subsequent problems with confident mapping of fusion sites within the transcripts [143]. The same is true also for the analysis of differential alternative splice sites [144].

At the next level, cancer RNA profiles can be compared with normal tissues to identify differentially expressed genes and calculate for them case-to-normal ratios [145,146]. The reference biomaterials can be typically obtained from either post-mortal healthy individuals or from surgically removed pathologically normal tissue adjacent to tumor [122]. Alternatively, many other normalization scenarios are possible, e.g. metastatic tissues vs corresponding primary tumors, tumors responding on treatment vs non-responder tumors, etc. [147,148]. The differential gene sets then can be analyzed in several ways.

First, case-to-normal expression ratios for the genes of a special attention can be interesting per se. Second, for the systemic analysis, the pools of upregulated and downregulated genes can be analyzed together of separately. This analysis may include interrogating enrichment of any specific Gene Ontology (GO) terms and their analogues [149,150] or can it deal with the quantitative analysis of functional features such as the activation of molecular pathways [128,151,152]. Remarkably, the latter type of analysis can be done not only on mRNA, but also on microRNA level [153].

All the metrics obtained from different levels of gene expression screenings in clinical biosamples (rough gene expression values, case-to-normal ratios, molecular pathway activation levels, etc) can be subjects of further applying *machine learning* (ML) methods to identify biomarkers or create robust classifiers. ML is defined as the study of algorithmically-built mathematical models that have been fitted for the portion of data called the *training dataset*, to make predictions for the similarly-obtained and similarly structured data called the *test* or *validation dataset* [154]. Efficiencies of ML-based models are described by specific quality metrics such as sensitivity (Sn), specificity (Sp), area under ROC curve (AUC), accuracy rate (ACC), Matthews correlation coefficient (MCC) or by p-values from statistical tests distinguishing one class from another [154–157]. Considering classical ML approaches,

most if not all of the available clinical genetic datasets are insufficient for solving the task of differentiating the classes [158–161] of features measured after RNA sequencing (e.g. gene expression values, case-to-normal ratios and even pathway activation levels) are far bigger than numbers of the corresponding patient biosamples linked with the clinical outcomes.

To generate statistically significant predictions, this requires extensive reduction of a pool of features to be considered, and ideally making their number comparable with the number of individuals analyzed [162]. This is the major limitation that can be only slightly attenuated by merging different clinical gene expression datasets. To reduce the number of features, they can be filtered according to specific functional of statistical traits (e.g. leaving only the genes coding for drug target genes; or genes/pathways with the highest abilities to discriminate treatment responders from non-responders in training datasets) [146]. The statistical methods for feature selection can be Pearson chi-squared test [163], correlation test [163], genetic algorithms [164], principal component analysis (PCA) [165], etc.

There is a selection of ML algorithms and their combinations to be further applied to a simplified expression dataset, like support vector machines (SVM) [166], *k* nearest neighbors [167], decision trees [163], random forest [165] and other methods. The data are initially obtained with the training dataset are then validated using independent validation dataset.

However, the demonstrated performance of standard ML classifiers was not high for clinically relevant predictions like drug response in cancer patients [166,168]. To address this challenge, a new paradigm recently emerged of considering flexible rather than fixed sets of features that are fitted individually for every particular comparison of a biosample with the pool of controls/training datasets [146,169]. The approach proposed utilized *data trimming* – sample-specific removal of features that don't have significant number of neighboring hits in the training dataset. Flexible data trimming prevents ML classifier from extrapolation by excluding too variable features. In a pilot application for the SVM method of ML and transcriptomic data, this enabled to dramatically increase number and quality of biomarkers predicting responses to chemotherapy treatments for 10/10 cohorts of 46–235 cancer patients [146].

12. Applications to clinical oncology

In its clinical oncology applications, RNA sequencing can be used alone or in combination with DNA mutation analysis. Using both types of data for high throughput genetic tests may combine strengths of mutation profiling and gene expression data. Mutations are powerful predictive biomarkers of several cancer drugs and therapeutic regimens [170,171]. So far, mutations are typically profiled using genomic DNA isolated from tumors, but not with RNA sequencing reads [172,173]. This is most probably due to frequently degraded RNAs in clinical samples, short sequencing reads and disproportionate coverage of different gene sequences because of drastically variable expression levels. Tumor mutation burden per million base pairs, which is one of the major genetic markers for selection of immunotherapy treatments, is also standardly measured using genetic (complete exome or clinical exome panel) data [174,175]. However, theoretically this type of analysis could be performed using RNA sequencing data as well, the major requirement here is having sufficiently covered several million base pairs of gene exon sequence.

In turn, RNA sequencing data make it possible to identify differentially expressed drug target genes [176] and measure activation of molecular pathways [151]. This allows patient-oriented personalized ranking of all potentially beneficial cancer drugs with known molecular specificities [151,152,177].

Immunohistochemistry remains a method of choice for interrogating expression of cancer markers in most of clinical laboratories [178–180]. However, RNA sequencing may provide even more accurate way of measuring expression of marker genes, as this is the case for *PDL1* gene which expression positively correlates with patient's response on anti-PD-1/PD-L1 immunotherapy [181].

In recent works it was found experimentally [102,182] and theoretically formulated [99,128] that RNA data-based molecular pathway activation levels outperform single gene expression levels as high-quality cancer biomarkers. Alternatively, single gene expression profiles can be aggregated in specific gene signature with specific overall score calculated on the basis of expression of relevant genes (see below) [183,184].

13. Gene expression signatures as classification and prognostic tools

Application of transcriptomic profiling to tumor classifications yielded clinically important results in breast cancer, where four distinct subtypes were defined in close concordance with major IHC markers (estrogen receptors, HER2 and Ki-67) [185]. Similar efforts in other cancers were less straightforward: e.g., for colorectal cancer, consensus molecular subtypes were defined only recently, and their clinical relevance, as well as reproducibility, remains to be elusive [186,187].

In addition, several gene expression signatures were developed to guide treatment (foremost, adjuvant chemotherapy) in breast cancer. MammaPrint, the 70-gene panel, was developed in 2002 and was tested in a large randomized controlled trial (MINDACT), confirming its prognostic power to differentiate low- and high-risk patients [188,189]. Along with MammaPrint, other transcriptomic signatures, namely, Oncotype Dx (21-gene signature recently tested in TAILORx trial), Prosigna (PAM 50, a 50-gene panel), and Endopredict (a 12-gene panel) received a regulatory approval [190–192]. While prognostic value of these signatures is well documented, predictive value in less certain (i.e., they can predict tumor relapse, but not whether chemotherapy can indeed prevent the dire outcome).

In turn, in thyroid cancer diagnostics currently microRNA molecule expression signatures are used. The major clinical problem here is the proper early diagnostics of thyroid cancer in thin needle biopsy samples. The cytological tests routinely in use for that often fail in discriminating follicular adenoma (benign neoplasm) and malignant follicular subtype of thyroid cancer. This frequently results in incorrect diagnostics, thus leading to either unnecessary highly invasive surgery or overlooked tumor progression [193]. Several diagnostic panels have been published to discriminate between benign and malignant thyroid neoplasms, all of them with the major impact of microRNA expression component [194–196]. For example, the marketed diagnostic panels RosettaGXreveal (Rosetta Genomics) and ThyraMIR (Interpace Diagnostic) are based on 24-microRNA and 10-microRNA (plus eight DNA mutations) expression signatures, respectively (reviewed in [194]).

Finally, there are several studies where transcriptomic data were used to adjust tumor therapies, all of them published very recently. In a multicenter trial termed WINTHER (NCT01856296) that started in 2013 and was published in 2019 [197], advanced cancer patients (the median number of previous therapies was three) underwent molecular diagnostics procedure. The patients were first investigated by DNA sequencing of their fresh biopsies using Foundation One clinical panel involving 236 genes exons. In part because of a small panel size, many tumors had no tractable genomic alterations. These who could not obtain meaningful results from DNA assay were subjected to transcriptome profiling using Agilent microarrays to establish expression levels of drug target genes.

The patients were then treated in accordance to DNA profiling (arm A) or RNA expression (arm B). In arm B, gene expression in tumor or metastases was compared with adjacent normal tissues. The clinical management committee (investigators from five countries) recommended therapies, prioritizing genomic matches; physicians determined the therapy given. Matching scores were calculated post-hoc for each patient, according to drugs received: for DNA, the number of

alterations matched divided by the total alteration number; for RNA, expression-matched drug ranks. More specifically, the method used for RNA data was based on a specific knowledge database comprising the target genes for each drug under consideration. Determining a score for each drug relied on the percentage of deregulated genes among the target genes implicated in the efficacy of each drug in the tumor sample from the patient.

Overall, 303 patients were included in this study; only 107 of them (35%; 69 in arm A and 38 in arm B) were evaluable for therapy. The most common diagnoses were colon, head and neck, and lung cancers. Among the 107 patients, the rate of stable disease exceeding 6 months and partial or complete response was 23.2% for arm A and 31.6% for arm B, thus showing somewhat better results for the RNA-based diagnostic cohort [197].

Alternative and more sophisticated transcriptome drug scoring approach termed Oncobox [198] is based on the analysis of intracellular molecular pathways activation and measuring expressions of molecular target genes for every anticancer target drug under consideration. Using Oncobox method requires collection of normal (control) expression profiles and annotated databases of molecular pathways and drug target genes. Both microarray and RNA sequencing data are acceptable for this type of analysis, although the latter type of data prevails in the most recent applications of this technique [198]. The method output is the personalized rating of target anticancer drugs prioritized according to *balanced efficiency score* (BES) that is calculated for every drug considering up/downregulation of its molecular targets and over/underactivation of relevant molecular pathways.

Previously, a pilot prospective clinical investigation was performed for a cohort of 23 recurrent/metastatic solid tumor patients using microarray gene expression data [145,199]. The objective response rate for the Oncobox-guided drug prescriptions was ˜ 61% (complete + partial response, RECIST). Since April 2018, a new trial started using RNA sequencing data for recurrent/metastatic solid tumors that includes 239 patients (trial ID NCT03724097). RNAs were extracted from the FFPE tissue blocks of surgically removed tumors or tumor biopsies. Preference was given to the most recently obtained tissue specimens. Following the test, 130 anticancer target drugs were rated according to their predicted effectiveness. After the appointment of therapy, the patients were naturally divided into the three observation groups. The first group was formed by patients receiving drugs in agreement with the Oncobox drug efficiency prediction as monotherapy or in combinations. In the second group, patients received drugs not recommended according to the Oncobox tests; in the third group patients received palliative care. This trial is ongoing, but several preliminary results have been recently published [200]. The primary feedback information was received for 144 patients. 25 patients (17%) died before prescription of the therapy, 19% received palliative care treatment, 39% received Oncobox-recommended therapies and 25% received other therapies. Tumor responses were estimated for 30 patients receiving therapies, with the control-over-disease rates of 71% for Oncobox-recommended and 44% for other therapies. The results obtained also suggested that cancer metastases and primary tumors frequently have different gene expression, intracellular molecular pathway activation and drug scoring patterns, thus pointing on the importance of testing multiple tumor sites.

The latter finding was in line with another published study using the same platform [201] where multisampling was used to develop a molecular guided tool for individualized selection of chemotherapeutics in recurrent glioblastoma (GBM). From 2016 to 2018, biopsies from primary and recurrent GBM were collected from 44 GBMs including 23 primary, 19 recurrent and 2 secondary recurrent cases. In parallel, biopsy materials were used to establish GBM stem cell lines. Both biopsy materials and cell cultures were examined by RNA sequencing and Oncobox analysis. Totally, 128 tissue samples and 42 cell cultures were investigated, for fourteen patients matching pairs of primary and recurrent GBM could be obtained. The results were compared for primary and recurrent GBM. Oncobox analysis showed downregulation of several pathways related to cell cycle and DNA repair and upregulation of pathways involved in immune response in recurrent GBM compared with the corresponding primary tumors. Specifically, in recurrent GBM there was a clear-cut down regulation of pathways targeted by previously administered chemotherapeutic Temozolomide. However, several other pathways were upregulated, including those targeted by drugs Durvalumab and Pomalidomide currently investigated in phase II or III trials for GBM. These results were similar for both tissue and cell culture specimens. Conclusion was drawn that the RNA sequencing information linked with the bioinformatic analysis using Oncobox platform has the potential of predicting sensitivity to chemotherapeutics in GBM on an individual basis. In addition, a significant degree of intratumoral heterogeneity (comparable to intertumoral heterogeneity) was detected in most of the GBM samples [201].

Finally, there are two recently published case report studies where Oncobox system was used to predict efficiencies of tyrosine kinase inhibitor (TKI) drugs in advanced metastatic cancers.

A 26-year-old woman with progressive granulosa cell ovarian tumor underwent salvage therapy selection despite multiple previous lines of therapy. The following target drugs were on the top of the rating according the Oncobox test: Regorafenib, Sorafenib, Sunitinib, Pazopanib, Axitinib, Imatinib. In October 2015, patient received treatment with Sorafenib, but it was not well tolerated and this therapy was terminated after 2 months. However, ultrasound examination indicated an associated decrease in the size for 3 out of 4 neoplasms. The therapy regimen was then switched to Imatinib, which was well tolerated and resulted in a disease stabilization (RECIST). As for February 2019, Imatinib administration was continued and patient is physically active with Karnofsky scale index 90% [202]. Another patient with progressive metastatic cholangiocarcinoma was prescribed with subsequent TKI monotherapies for Sorafenib and Pazopanib that occupied, respectively, 2nd and 4th positions of the Oncobox personalized drug ratings. It resulted in a two-years stabilization (RECIST) [203].

In principle, similar pathway approaches that were effective for single drugs may also work for their combinations. However, these studies are now at an early stage with only few recent reports publishing finding effective ATD combinations for cancer cell line or animal models [204–207].

14. Concluding remarks

In the previous sections of this review, we tried to address the current progress in technology, concepts and clinical applications of RNA sequencing in oncology. Several lines of evidence suggest that molecular tests based on transcriptomics are more efficient when combined with DNA mutation analysis (see above). In a recent publication by Zolotovskaya et al., a method was published pioneering quantitative molecular pathway analysis based on complete cancer exome sequencing profiles [145,208,209]. A metric termed “pathway instability” (PI) was introduced that reflects overall mutation burden of a pathway. It can be calculated for either total mutations or any specific group of them such as truncating mutations that abrogate protein functions [209]. Like in transcriptomic data metrics, PI serves as a significantly better type of biomarkers compared to mutations in individual genes [209]. Furthermore, a pathway-based algorithm was developed using PI values to predict clinical efficacies of drugs. The output value termed Mutation Drug Scoring (MDS) positively correlated with the expected efficacies of drugs for specific tumors, as investigated using 3.800 exome mutation profiles for 128 drugs by finding correlations of MDS values with the known drug efficiencies from clinical trials [208]. We may expect further intensive development of this field, most probably resulting in an effective symbiosis of RNA- and DNA screening methods within next generation cancer diagnostic platforms (Fig.2).

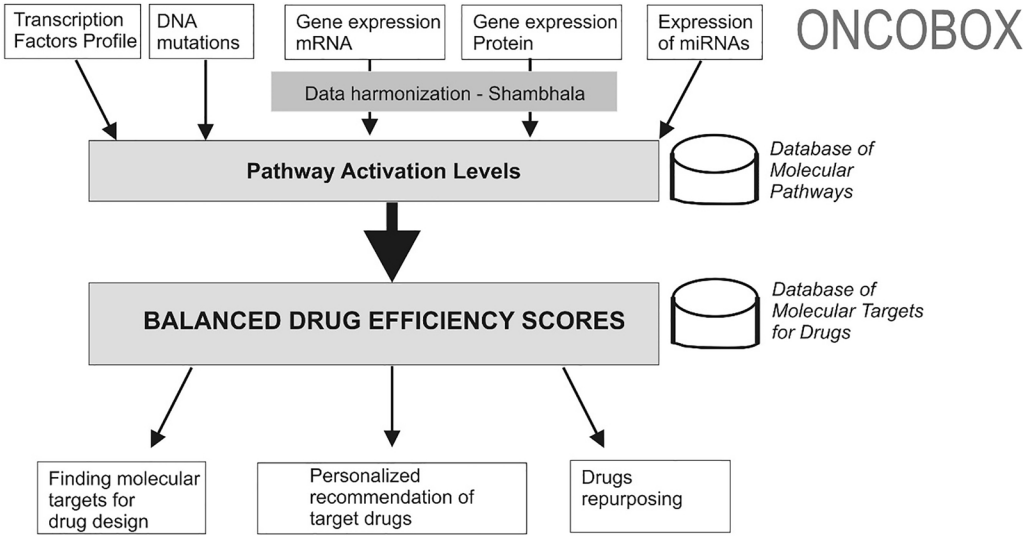

Fig. 2. Example of a platform integrating multi-OMICS data for personalized oncology and drug development/repurposing [198].

Declaration of Competing Interest

The authors declare that they have no conflict of interests in publishing this review paper

Acknowledgements

The study was supported by Russian Foundation for Basic Research Grant 19-29-01108. We thank Prof. Ilya Muchnik (Rutgers University, USA), Prof. Nicolas Borisov (Moscow Sechenov Medical University, Russia) and Prof. Donald Geman (Johns Hopkins University, USA) for insightful discussion.

References

[1] D.G. Maloney, A.J. Grillo-López, C.A. White, D. Bodkin, R.J. Schilder, J.A. Neidhart, N. Janakiraman, K.A. Foon, T.M. Liles, B.K. Dallaire, K. Wey, I. Royston, T. Davis, R. Levy, IDEC-C2B8 (Rituximab) anti-CD20 monoclonal antibody therapy in patients with relapsed low-grade non-Hodgkin's lymphoma, Blood. 90 (1997) 2188–2195 (accessed June 26, 2019), http://www.ncbi.nlm.nih.gov/pubmed/9310469.

[2] B.J. Druker, M. Talpaz, D.J. Resta, B. Peng, E. Buchdunger, J.M. Ford, N.B. Lydon, H. Kantarjian, R. Capdeville, S. Ohno-Jones, C.L. Sawyers, Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia, N. Engl. J. Med. 344 (2001) 1031–1037, https://doi.org/10.1056/NEJM200104053441401.

[3] D.J. Slamon, B. Leyland-Jones, S. Shak, H. Fuchs, V. Paton, A. Bajamonde, T. Fleming, W. Eiermann, J. Wolter, M. Pegram, J. Baselga, L. Norton, Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2, N. Engl. J. Med. 344 (2001) 783–792, https://doi.org/10.1056/NEJM200103153441101.

[4] P. Vihinen, R. Ala-aho, V.-M. Kähäri, Matrix metalloproteinases as therapeutic targets in cancer, Curr. Cancer Drug Targets 5 (2005) 203–220 (accessed June 26, 2019), http://www.ncbi.nlm.nih.gov/pubmed/15892620.

[5] F. Ciardiello, G. Tortora, EGFR antagonists in cancer treatment, N. Engl. J. Med. 358 (2008) 1160–1174, https://doi.org/10.1056/NEJMra0707704.

[6] J.M.L. Ostrem, K.M. Shokat, Direct small-molecule inhibitors of KRAS: from structural insights to mechanism-based design, Nat. Rev. Drug Discov. 15 (2016) 771–785, https://doi.org/10.1038/nrd.2016.139.

[7] G. Gasparini, R. Longo, M. Fanelli, B.A. Teicher, Combination of antiangiogenic therapy with other anticancer therapies: results, challenges, and Open questions, J. Clin. Oncol. 23 (2005) 1295–1311, https://doi.org/10.1200/JCO.2005.10.022.

[8] N. Ferrara, R.S. Kerbel, Angiogenesis as a therapeutic target, Nature. 438 (2005) 967–974, https://doi.org/10.1038/nature04483.

[9] T.J. Lynch, D.W. Bell, R. Sordella, S. Gurubhagavatula, R.A. Okimoto, B.W. Brannigan, P.L. Harris, S.M. Haserlat, J.G. Supko, F.G. Haluska, D.N. Louis, D.C. Christiani, J. Settleman, D.A. Haber, Activating mutations in the epidermal growth factor receptor underlying responsiveness of Non–Small-cell lung cancer to gefitinib, N. Engl. J. Med. 350 (2004) 2129–2139, https://doi.org/10.1056/NEJMoa040938.

[10] C. Gridelli, A. Rossi, D.P. Carbone, J. Guarize, N. Karachaliou, T. Mok, F. Petrella, L. Spaggiari, R. Rosell, Non-small-cell lung cancer, Nat. Rev. Dis. Prim. 1 (2015) 15009, https://doi.org/10.1038/nrdp.2015.9.

[11] M. Reck, K.F. Rabe, Precision diagnosis and treatment for advanced Non-small-cell lung cancer, N. Engl. J. Med. 377 (2017) 849–861, https://doi.org/10.1056/NEJMra1703413.

[12] N. Normanno, S. Tejpar, F. Morgillo, A. De Luca, E. Van Cutsem, F. Ciardiello, Implications for KRAS status and EGFR-targeted therapies in metastatic CRC, Nat. Rev. Clin. Oncol. 6 (2009) 519–527, https://doi.org/10.1038/nrclinonc.2009.111.

[13] A. Lièvre, J.-B. Bachet, D. Le Corre, V. Boige, B. Landi, J.-F. Emile, J.-F. Côté, G. Tomasic, C. Penna, M. Ducreux, P. Rougier, F. Penault-Llorca, P. Laurent-Puig, KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer, Cancer Res. 66 (2006) 3992–3995, https://doi.org/10.1158/0008-5472.CAN-06-0191.

[14] K.T. Flaherty, I. Puzanov, K.B. Kim, A. Ribas, G.A. McArthur, J.A. Sosman, P.J. O'Dwyer, R.J. Lee, J.F. Grippo, K. Nolop, P.B. Chapman, Inhibition of mutated, activated BRAF in metastatic melanoma, N. Engl. J. Med. 363 (2010) 809–819, https://doi.org/10.1056/NEJMoa1002011.

[15] C. Robert, B. Karaszewska, J. Schachter, P. Rutkowski, A. Mackiewicz, D. Stroiakovski, M. Lichinitser, R. Dummer, F. Grange, L. Mortier, V. Chiarion-Sileni, K. Drucis, I. Krajsova, A. Hauschild, P. Lorigan, P. Wolter, G.V. Long, K. Flaherty, P. Nathan, A. Ribas, A.-M. Martin, P. Sun, W. Crist, J. Legos, S.D. Rubin, S.M. Little, D. Schadendorf, Improved overall survival in melanoma with combined dabrafenib and trametinib, N. Engl. J. Med. 372 (2015) 30–39, https://doi.org/10.1056/NEJMoa1412690.

[16] D. Planchard, B. Besse, H.J.M. Groen, P.-J. Souquet, E. Quoix, C.S. Baik, F. Barlesi, T.M. Kim, J. Mazieres, S. Novello, J.R. Rigas, A. Upalawanna, A.M. D'Amelio, P. Zhang, B. Mookerjee, B.E. Johnson, Dabrafenib plus trametinib in patients with previously treated BRAF(V600E)-mutant metastatic non-small cell lung cancer: an open-label, multicentre phase 2 trial, Lancet. Oncol. 17 (2016) 984–993, https://doi.org/10.1016/S1470-2045(16)30146-2.

[17] Z. Karoulia, E. Gavathiotis, P.I. Poulikakos, New perspectives for targeting RAF kinase in human cancer, Nat. Rev. Cancer. 17 (2017) 676–691, https://doi.org/10.1038/nrc.2017.79.

[18] E. Sanz-Garcia, G. Argiles, E. Elez, J. Tabernero, BRAF mutant colorectal cancer: prognosis, treatment, and new perspectives, Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. 28 (2017) 2648–2657, https://doi.org/10.1093/annonc/mdx401.

[19] M. Soda, Y.L. Choi, M. Enomoto, S. Takada, Y. Yamashita, S. Ishikawa, S. Fujiwara, H. Watanabe, K. Kurashina, H. Hatanaka, M. Bando, S. Ohno, Y. Ishikawa, H. Aburatani, T. Niki, Y. Sohara, Y. Sugiyama, H. Mano, Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer, Nature. 448 (2007) 561–566, https://doi.org/10.1038/nature05945.

[20] E.L. Kwak, Y.-J. Bang, D.R. Camidge, A.T. Shaw, B. Solomon, R.G. Maki, Ou S-HI, B.J. Dezube, P.A. Jänne, D.B. Costa, M. Varella-Garcia, W.-H. Kim, T.J. Lynch, P. Fidias, H. Stubbs, J.A. Engelman, L.V. Sequist, W. Tan, L. Gandhi, M. Mino-Kenudson, G.C. Wei, S.M. Shreeve, M.J. Ratain, J. Settleman, J.G. Christensen, D.A. Haber, K. Wilner, R. Salgia, G.I. Shapiro, J.W. Clark, A.J. Iafrate, Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer, N. Engl. J. Med. 363 (2010) 1693–1703, https://doi.org/10.1056/NEJMoa1006448.

[21] D.D. Von Hoff, P.M. LoRusso, C.M. Rudin, J.C. Reddy, R.L. Yauch, R. Tibes, G.J. Weiss, M.J. Borad, C.L. Hann, J.R. Brahmer, H.M. Mackey, B.L. Lum, W.C. Darbonne, J.C. Marsters, F.J. de Sauvage, J.A. Low, Inhibition of the hedgehog pathway in advanced basal-cell carcinoma, N. Engl. J. Med. 361 (2009) 1164–1172, https://doi.org/10.1056/NEJMoa0905360.

[22] G.W. Robinson, B.A. Orr, G. Wu, S. Gururangan, T. Lin, I. Qaddoumi, R.J. Packer, S. Goldman, M.D. Prados, A. Desjardins, M. Chintagumpala, N. Takebe, S.C. Kaste, M. Rusch, S.J. Allen, A. Onar-Thomas, C.F. Stewart, M. Fouladi, J.M. Boyett, R.J. Gilbertson, T. Curran, D.W. Ellison, A. Gajjar, Vismodegib Exerts targeted

efficacy against recurrent sonic hedgehog-subgroup medulloblastoma: results from phase II pediatric brain tumor consortium studies PBTC-025B and PBTC-032, J. Clin. Oncol. 33 (2015) 2646–2654, https://doi.org/10.1200/JCO.2014.60.1591.

[23] A.T. Shaw, Ou S-HI, Y.-J. Bang, D.R. Camidge, B.J. Solomon, R. Salgia, G.J. Riely, M. Varella-Garcia, G.I. Shapiro, D.B. Costa, R.C. Doebele, Zheng Z. Le LP, W. Tan, P. Stephenson, S.M. Shreeve, L.M. Tye, J.G. Christensen, K.D. Wilner, J.W. Clark, A.J. Iafrate, Crizotinib in ROS1-rearranged non-small-cell lung cancer, N. Engl. J. Med. 371 (2014) 1963–1971, https://doi.org/10.1056/NEJMoa1406766.

[24] A. Drilon, T.W. Laetsch, S. Kummar, S.G. DuBois, U.N. Lassen, G.D. Demetri, M. Nathenson, R.C. Doebele, A.F. Farago, A.S. Pappo, B. Turpin, A. Dowlati, M.S. Brose, L. Mascarenhas, N. Federman, J. Berlin, W.S. El-Deiry, C. Baik, J. Deeken, V. Boni, R. Nagasubramanian, M. Taylor, E.R. Rudzinski, F. Meric-Bernstam, D.P.S. Sohal, P.C. Ma, L.E. Raez, J.F. Hechtman, R. Benayed, M. Ladanyi, B.B. Tuch, K. Ebata, S. Cruickshank, N.C. Ku, M.C. Cox, D.S. Hawkins, D.S. Hong, D.M. Hyman, Efficacy of larotrectinib in TRK fusion-positive cancers in adults and children, N. Engl. J. Med. 378 (2018) 731–739, https://doi.org/10.1056/NEJMoa1714448.

[25] P.K. Paik, A. Drilon, P.-D. Fan, H. Yu, N. Rekhtman, M.S. Ginsberg, L. Borsu, N. Schultz, M.F. Berger, C.M. Rudin, M. Ladanyi, Response to MET inhibitors in patients with stage IV lung adenocarcinomas harboring MET mutations causing exon 14 skipping, Cancer Discov. 5 (2015) 842–849, https://doi.org/10.1158/2159-8290.CD-14-1467.

[26] C.J. Lord, A. Ashworth, PARP inhibitors: synthetic lethality in the clinic, Science. 355 (2017) 1152–1158, https://doi.org/10.1126/science.aam7344.

[27] G.T. Gibney, L.M. Weiner, M.B. Atkins, Predictive biomarkers for checkpoint inhibitor-based immunotherapy, Lancet. Oncol. 17 (2016) e542–e551, https://doi.org/10.1016/S1470-2045(16)30406-5.

[28] D.R. Camidge, R.C. Doebele, K.M. Kerr, Comparing and contrasting predictive biomarkers for immunotherapy and targeted therapy of NSCLC, Nat. Rev. Clin. Oncol. 16 (2019) 341–355, https://doi.org/10.1038/s41571-019-0173-9.

[29] J.J. Havel, D. Chowell, T.A. Chan, The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy, Nat. Rev. Cancer. 19 (2019) 133–150, https://doi.org/10.1038/s41568-019-0116-x.

[30] D.T. Le, J.N. Uram, H. Wang, B.R. Bartlett, H. Kemberling, A.D. Eyring, A.D. Skora, B.S. Luber, N.S. Azad, D. Laheru, B. Biedrzycki, R.C. Donehower, A. Zaheer, G.A. Fisher, T.S. Crocenzi, J.J. Lee, S.M. Duffy, R.M. Goldberg, A. de la Chapelle, M. Koshiji, F. Bhaijee, T. Huebner, R.H. Hruban, L.D. Wood, N. Cuka, D.M. Pardoll, N. Papadopoulos, K.W. Kinzler, S. Zhou, T.C. Cornish, J.M. Taube, R.A. Anders, J.R. Eshleman, B. Vogelstein, L.A. Diaz, PD-1 blockade in tumors with mismatch-repair deficiency, N. Engl. J. Med. 372 (2015) 2509–2520, https://doi.org/10.1056/NEJMoa1500596.

[31] K.W. Hunter, R. Amin, S. Deasy, N.-H. Ha, L. Wakefield, Genetic insights into the morass of metastatic heterogeneity, Nat. Rev. Cancer. 18 (2018) 211–223, https://doi.org/10.1038/nrc.2017.126.

[32] A. Popovtzer, M. Sarfaty, D. Limon, G. Marshack, E. Perlow, A. Dvir, L. Soussan-Gutman, S.M. Stemmer, Metastatic salivary gland tumors: A single-center study demonstrating the feasibility and potential clinical benefit of molecular-profiling-guided therapy, Biomed. Res. Int. 2015 (2015) 614845, , https://doi.org/10.1155/2015/614845.

[33] J. Vigneswaran, Tan Y-HC, S.D. Murgu, B.M. Won, K.A. Patton, V.M. Villaflor, P.C. Hoffman, T. Hensing, D.K. Hogarth, R. Malik, H. MacMahon, J. Mueller, C.A. Simon, W.T. Vigneswaran, C.H. Wigfield, M.K. Ferguson, A.N. Husain, E.E. Vokes, R. Salgia, Comprehensive genetic testing identifies targetable genomic alterations in most patients with non-small cell lung cancer, specifically adenocarcinoma, single institute investigation, Oncotarget. 7 (2016) 18876–18886, https://doi.org/10.18632/oncotarget.7739.

[34] B.K. Neilsen, R. Sleightholm, R. McComb, S.H. Ramkissoon, J.S. Ross, R.J. Corona, V.A. Miller, M. Cooke, M.R. Aizenberg, Comprehensive genetic alteration profiling in primary and recurrent glioblastoma, J. Neurooncol. 142 (2019) 111–118, https://doi.org/10.1007/s11060-018-03070-2.

[35] K.V. Ruggles, K. Krug, X. Wang, K.R. Clauser, J. Wang, S.H. Payne, D. Fenyö, B. Zhang, D.R. Mani, Methods, tools and current perspectives in proteogenomics, Mol. Cell. Proteom.. 16 (2017) 959–981, https://doi.org/10.1074/mcp.MR117.000024.

[36] J.T. Painter, N.P. Clayton, R.A. Herbert, Useful immunohistochemical markers of tumor differentiation, Toxicol. Pathol. 38 (2010) 131–141, https://doi.org/10.1177/0192623309356449.

[37] L. Stephen, Multiplex immunoassay profiling, Methods Mol. Biol. 1546 (2017) 169–176, https://doi.org/10.1007/978-1-4939-6730-8_13.

[38] J.G. Duarte, J.M. Blackburn, Advances in the development of human protein microarrays, Expert Rev. Proteom.. 14 (2017) 627–641, https://doi.org/10.1080/14789450.2017.1347042.

[39] R. Aebersold, M. Mann, Mass-spectrometric exploration of proteome structure and function, Nature. 537 (2016) 347–355, https://doi.org/10.1038/nature19949.

[40] J.W. Lee, D. Figeys, J. Vasilescu, Biomarker assay translation from discovery to clinical studies in cancer drug development: quantification of emerging protein biomarkers, Adv. Cancer Res. 96 (2007) 269–298, https://doi.org/10.1016/S0065-230X(06)96010-2.

[41] T.W. Chang, Binding of cells to matrixes of distinct antibodies coated on solid surface, J. Immunol. Methods 65 (1983) 217–223 (accessed June 26, 2019), http://www.ncbi.nlm.nih.gov/pubmed/6606681.

[42] A. Sreekumar, M.K. Nyati, S. Varambally, T.R. Barrette, D. Ghosh, T.S. Lawrence, A.M. Chinnaiyan, Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins, Cancer Res. 61 (2001) 7585–7593 (accessed June 26, 2019), http://www.ncbi.nlm.nih.gov/pubmed/11606398.

[43] J.E. Mirus, Y. Zhang, M.A. Hollingsworth, J.L. Solan, P.D. Lampe, S.R. Hingorani, Spatiotemporal proteomic analyses during pancreas cancer progression identifies serine/threonine stress kinase 4 (STK4) as a novel candidate biomarker for early stage disease, Mol. Cell. Proteom.. 13 (2014) 3484–3496, https://doi.org/10.1074/mcp.M113.036517.

[44] L.C. Stetson, J.-E. Dazard, J.S. Barnholtz-Sloan, Protein markers predict survival in glioma patients, Mol. Cell. Proteom.. 15 (2016) 2356–2365, https://doi.org/10.1074/mcp.M116.060657.

[45] M. Witt, J.-G. Walter, Stahl F. Aptamer microarrays-current Status and future prospects, Microarrays (Basel, Switz.). 4 (2015) 115–132, https://doi.org/10.3390/microarrays4020115.

[46] J. Webber, T.C. Stone, E. Katilius, B.C. Smith, B. Gordon, M.D. Mason, Z. Tabi, I.A. Brewis, A. Clayton, Proteomics analysis of cancer exosomes using a novel modified aptamer-based array (SOMAscan™) platform, Mol. Cell. Proteom.. 13 (2014) 1050–1064, https://doi.org/10.1074/mcp.M113.032136.

[47] E.N. Brody, L. Gold, R.M. Lawn, J.J. Walker, D. Zichi, High-content affinity-based proteomics: unlocking protein biomarker discovery, Expert Rev. Mol. Diagn. 10 (2010) 1013–1022, https://doi.org/10.1586/erm.10.89.

[48] J. Duo, C. Chiriac, Huang RY-C, J. Mehl, G. Chen, A. Tymiak, P. Sabbatini, R. Pillutla, Y. Zhang, Slow off-rate modified aptamer (SOMAmer) as a novel reagent in immunoassay development for accurate soluble glypican-3 quantification in clinical samples, Anal. Chem. 90 (2018) 5162–5170, https://doi.org/10.1021/acs.analchem.7b05277.

[49] K.M. Bhawe, M.K. Aghi, Microarray analysis in glioblastomas, Methods Mol. Biol. 1375 (2016) 195–206, https://doi.org/10.1007/7651_2015_245.

[50] J.M. Rosenberg, P.J. Utz, Protein microarrays: a new tool for the study of autoantibodies in immunodeficiency, Front. Immunol. 6 (2015) 138, https://doi.org/10.3389/fimmu.2015.00138.

[51] M. Mann, Origins of mass spectrometry-based proteomics, Nat. Rev. Mol. Cell. Biol. 17 (2016) 678, https://doi.org/10.1038/nrm.2016.135.

[52] M.A. Karpova, S.A. Moshkovskii, I.Y. Toropygin, A.I. Archakov, Cancer-specific MALDI-TOF profiles of blood serum and plasma: biological meaning and perspectives, J. Proteom.. 73 (2010) 537–551, https://doi.org/10.1016/j.jprot.2009.09.011.

[53] E.F. Petricoin, A.M. Ardekani, B.A. Hitt, P.J. Levine, V.A. Fusaro, S.M. Steinberg, G.B. Mills, C. Simone, D.A. Fishman, E.C. Kohn, L.A. Liotta, Use of proteomic patterns in serum to identify ovarian cancer, Lancet (Lond., Engl.). 359 (2002) 572–577, https://doi.org/10.1016/S0140-6736(02)07746-2.

[54] S.A. Moshkovskii, M.A. Vlasova, M.A. Pyatnitskiy, O.V. Tikhonova, M.R. Safarova, O.V. Makarov, A.I. Archakov, Acute phase serum amyloid A in ovarian cancer as an important component of proteome diagnostic profiling, Proteom.. Clin. Appl. 1 (2007) 107–117, https://doi.org/10.1002/prca.200600229.

[55] Z. Zhang, R.C. Bast, Y. Yu, J. Li, L.J. Sokoll, A.J. Rai, J.M. Rosenzweig, B. Cameron, Y.Y. Wang, X.-Y. Meng, A. Berchuck, C. van Haaften-Day, N.F. Hacker, H.W.A. de Bruijn, A.G.J. van der Zee, I.J. Jacobs, E.T. Fung, D.W. Chan, Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer, Cancer Res. 64 (2004) 5882–5890, https://doi.org/10.1158/0008-5472.CAN-04-0746.

[56] M.J. Fidler, C.L. Fhied, J. Roder, S. Basu, S. Sayidine, I. Fughhi, M. Pool, M. Batus, P. Bonomi, J.A. Borgia, The serum-based VeriStrat® test is associated with proinflammatory reactants and clinical outcome in non-small cell lung cancer patients, BMC Cancer. 18 (2018) 310, https://doi.org/10.1186/s12885-018-4193-0.

[57] F. Coscia, K.M. Watters, M. Curtis, M.A. Eckert, C.Y. Chiang, S. Tyanova, A. Montag, R.R. Lastra, E. Lengyel, M. Mann, Integrative proteomic profiling of ovarian cancer cell lines reveals precursor cell associated proteins and functional status, Nat. Commun. 7 (2016) 12645, https://doi.org/10.1038/ncomms12645.

[58] L.H. Betancourt, K. Pawłowski, J. Eriksson, A.M. Szasz, S. Mitra, I. Pla, C. Welinder, H. Ekedahl, P. Broberg, R. Appelqvist, M. Yakovleva, Y. Sugihara, K. Miharada, C. Ingvar, L. Lundgren, B. Baldetorp, H. Olsson, M. Rezeli, E. Wieslander, P. Horvatovich, J. Malm, G. Jönsson, G. Marko-Varga, Improved survival prognostication of node-positive malignant melanoma patients utilizing shotgun proteomics guided by histopathological characterization and genomic data, Sci. Rep. 9 (2019) 5154, https://doi.org/10.1038/s41598-019-41625-z.

[59] S. Principe, S. Mejia-Guerrero, V. Ignatchenko, A. Sinha, A. Ignatchenko, W. Shi, K. Pereira, S. Su, S.H. Huang, B. O'Sullivan, W. Xu, D.P. Goldstein, I. Weinreb, L. Ailles, F.-F. Liu, T. Kislinger, Proteomic analysis of cancer-associated fibroblasts reveals a paracrine role for MFAP5 in human Oral Tongue squamous cell carcinoma, J. Proteome Res. 17 (2018) 2045–2059, https://doi.org/10.1021/acs.jproteome.7b00925.

[60] J.R. O'Neill, H.-S. Pak, E. Pairo-Castineira, V. Save, S. Paterson-Brown, R. Nenutil, B. Vojtěšek, I. Overton, A. Scherl, T.R. Hupp, Quantitative shotgun proteomics unveils candidate novel esophageal adenocarcinoma (EAC)-specific proteins, Mol. Cell. Proteom.. 16 (2017) 1138–1150, https://doi.org/10.1074/mcp.M116.065078.

[61] W. Yang, M.R. Freeman, N. Kyprianou, Personalization of prostate cancer therapy through phosphoproteomics, Nat. Rev. Urol. 15 (2018) 483–497, https://doi.org/10.1038/s41585-018-0014-0.

[62] B.M. Kuenzi, L.L. Remsing Rix, P.A. Stewart, B. Fang, F. Kinose, A.T. Bryant, T.A. Boyle, J.M. Koomen, E.B. Haura, U. Rix, Polypharmacology-based ceritinib repurposing using integrated functional proteomics, Nat. Chem. Biol. 13 (2017) 1222–1231, https://doi.org/10.1038/nchembio.2489.

[63] L. Giusti, C. Angeloni, A. Lucacchini, Update on proteomic studies of formalin-fixed paraffin-embedded tissues, Expert Rev. Proteom.. 16 (2019) 513–520, https://doi.org/10.1080/14789450.2019.1615452.

[64] M. Bayer, L. Angenendt, C. Schliemann, W. Hartmann, S. König, Are formalin-fixed and paraffin-embedded tissues fit for proteomic analysis? J. Mass. Spectrom.

(2019), https://doi.org/10.1002/jms.4347.
[65] R.W. Sprung, J.W.C. Brock, J.P. Tanksley, M. Li, M.K. Washington, R.J.C. Slebos, D.C. Liebler, Equivalence of protein inventories obtained from formalin-fixed paraffin-embedded and frozen tissue in multidimensional liquid chromatography-tandem mass spectrometry shotgun proteomic analysis, Mol. Cell. Proteom.. 8 (2009) 1988–1998, https://doi.org/10.1074/mcp.M800518-MCP200.
[66] B. Zhang, J. Wang, X. Wang, J. Zhu, Q. Liu, Z. Shi, M.C. Chambers, L.J. Zimmerman, K.F. Shaddox, S. Kim, S.R. Davies, S. Wang, P. Wang, C.R. Kinsinger, R.C. Rivers, H. Rodriguez, R.R. Townsend, M.J.C. Ellis, S.A. Carr, D.L. Tabb, R.J. Coffey, R.J.C. Slebos, D.C. Liebler, Proteogenomic characterization of human colon and rectal cancer, Nature. 513 (2014) 382–387, https://doi.org/10.1038/nature13438.
[67] S. Vasaikar, C. Huang, X. Wang, V.A. Petyuk, S.R. Savage, B. Wen, Y. Dou, M. Watson, et al., Proteogenomic analysis of human colon cancer reveals New therapeutic opportunities, Cell. 177 (2019) 1035–1049, https://doi.org/10.1016/j.cell.2019.03.030 e19.
[68] Y.J. Kim, S.M.M. Sweet, J.D. Egertson, A.J. Sedgewick, S. Woo, W.-L. Liao, G.E. Merrihew, B.C. Searle, C. Vaske, R. Heaton, M.J. MacCoss, T. Hembrough, Data-Independentindependent acquisition mass spectrometry to quantify protein levels in FFPE tumor biopsies for molecular diagnostics, J. Proteome Res. 18 (2019) 426–435, https://doi.org/10.1021/acs.jproteome.8b00699.
[69] C. Steiner, P. Lescuyer, J.-C. Tille, P. Cutler, A. Ducret, Development of a highly multiplexed SRM assay for biomarker Discovery in formalin-fixed paraffin-embedded tissues, Methods Mol. Biol. 1959 (2019) 185–203, https://doi.org/10.1007/978-1-4939-9164-8_13.
[70] F. Hoffmann, C. Umbreit, T. Krüger, D. Pelzel, G. Ernst, O. Kniemeyer, O. Guntinas-Lichius, A. Berndt, F. von Eggeling, Identification of proteomic markers in head and neck cancer using MALDI-MS imaging, LC-MS/MS, and immunohistochemistry, Proteom.. Clin. Appl. 13 (2019) e1700173, , https://doi.org/10.1002/prca.201700173.
[71] B. Zhang, J.R. Whiteaker, A.N. Hoofnagle, G.S. Baird, K.D. Rodland, A.G. Paulovich, Clinical potential of mass spectrometry-based proteogenomics, Nat. Rev. Clin. Oncol. 16 (2019) 256–268, https://doi.org/10.1038/s41571-018-0135-7.
[72] A. Polyakova, K. Kuznetsova, S. Moshkovskii, Proteogenomics meets cancer immunology: mass spectrometric discovery and analysis of neoantigens, Expert Rev. Proteom. 12 (2015) 533–541, https://doi.org/10.1586/14789450.2015.1070100.
[73] P. Mertins, D.R. Mani, K.V. Ruggles, M.A. Gillette, K.R. Clauser, P. Wang, X. Wang, J.W. Qiao, S. Cao, F. Petralia, E. Kawaler, F. Mundt, K. Krug, Z. Tu, J.T. Lei, M.L. Gatza, M. Wilkerson, C.M. Perou, V. Yellapantula, K. Huang, C. Lin, M.D. McLellan, P. Yan, S.R. Davies, R.R. Townsend, S.J. Skates, J. Wang, B. Zhang, C.R. Kinsinger, M. Mesri, H. Rodriguez, L. Ding, A.G. Paulovich, D. Fenyö, M.J. Ellis, S.A. Carr, N.C.I. CPTAC, Proteogenomics connects somatic mutations to signalling in breast cancer, Nature. 534 (2016) 55–62, https://doi.org/10.1038/nature18003.
[74] H. Zhang, T. Liu, Z. Zhang, S.H. Payne, B. Zhang, J.E. McDermott, J.-Y. Zhou, R.R. Townsend, et al., Integrated proteogenomic characterization of human High-grade serous ovarian cancer, Cell. 166 (2016) 755–765, https://doi.org/10.1016/j.cell.2016.05.069.
[75] H.A. King, A.P. Gerber, Translatome profiling: methods for genome-scale analysis of mRNA translation, Brief. Funct. Genomics. 15 (2016) 22–31, https://doi.org/10.1093/bfgp/elu045.
[76] N.T. Ingolia, S. Ghaemmaghami, J.R.S. Newman, J.S. Weissman, Genome-Wide analysis in vivo of translation with nucleotide Resolution using ribosome profiling, Science (80-.) 324 (2009) 218–223, https://doi.org/10.1126/science.1168978.
[77] A.M. Michel, P.V. Baranov, Ribosome profiling: a Hi-def monitor for protein synthesis at the genome-wide scale, Wiley Interdiscip. Rev. RNA. 4 (2013) 473–490, https://doi.org/10.1002/wrna.1172.
[78] J. Zhao, B. Qin, R. Nikolay, C.M.T. Spahn, G. Zhang, Translatomics: The global View of translation, Int. J. Mol. Sci. 20 (2019) 212, https://doi.org/10.3390/ijms20010212.
[79] A. Bartholomäus, C. Del Campo, Z. Ignatova, Mapping the non-standardized biases of ribosome profiling, Biol. Chem. 397 (2016) 23–35, https://doi.org/10.1515/hsz-2015-0197.
[80] B. Liu, G. Molinaro, H. Shu, E.E. Stackpole, K.M. Huber, J.D. Richter, Optimization of ribosome profiling using low-input brain tissue from fragile X syndrome model mice, Nucleic Acids Res. 47 (2019) e25, https://doi.org/10.1093/nar/gky1292.
[81] L. Ma, Z. Liang, H. Zhou, L. Qu, Applications of RNA indexes for precision oncology in breast cancer, Genomics. Proteom. Bioinform.. 16 (2018) 108–119, https://doi.org/10.1016/j.gpb.2018.03.002.
[82] N. Bossel Ben-Moshe, S. Gilad, G. Perry, S. Benjamin, N. Balint-Lahat, A. Pavlovsky, S. Halperin, B. Markus, A. Yosepovich, I. Barshack, E.N. Gal-Yam, E. Domany, B. Kaufman, M. Dadiani, mRNA-seq whole transcriptome profiling of fresh frozen versus archived fixed tissues, BMC Genomics. 19 (2018) 419, https://doi.org/10.1186/s12864-018-4761-3.
[83] R. Edgar, M. Domrachev, A.E. Lash, Gene expression omnibus: NCBI gene expression and hybridization array data repository, Nucleic Acids Res. 30 (2002) 207 (accessed October 5, 2018), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC99122/.
[84] A. Brazma, H. Parkinson, U. Sarkans, M. Shojatalab, J. Vilo, N. Abeygunawardena, E. Holloway, M. Kapushesky, P. Kemmeren, G.G. Lara, A. Oezcimen, P. Rocca-Serra, S.-A. Sansone, ArrayExpress—a public repository for microarray gene expression data at the EBI, Nucleic Acids Res. 31 (2003) 68 (accessed October 5, 2018), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC165538/.
[85] T. Barrett, S.E. Wilhite, P. Ledoux, C. Evangelista, I.F. Kim, M. Tomashevsky, K.A. Marshall, K.H. Phillippy, P.M. Sherman, M. Holko, A. Yefanov, H. Lee, N. Zhang, C.L. Robertson, N. Serova, S. Davis, A. Soboleva, NCBI GEO: archive for functional genomics data sets—update, Nucleic Acids Res. 41 (2012) D991–D995, https://doi.org/10.1093/nar/gks1193.
[86] E. de Klerk, Fokkema IFAC, Thiadens KAMH, J.J. Goeman, M. Palmblad, J.T. den Dunnen, M. von Lindern, P.A.C. t Hoen, Assessing the translational landscape of myogenic differentiation by ribosome profiling, Nucleic Acids Res. 43 (2015) 4408, https://doi.org/10.1093/NAR/GKV281.
[87] K.C. Barry, N.T. Ingolia, R.E. Vance, Global analysis of gene expression reveals mRNA superinduction is required for the inducible immune response to a bacterial pathogen, Elife. 6 (2017), https://doi.org/10.7554/eLife.22707.
[88] J.G. Dunn, C.K. Foo, N.G. Belletier, E.R. Gavis, J.S. Weissman, Ribosome profiling reveals pervasive and regulated stop codon readthrough in drosophila melanogaster, Elife. 2 (2013) e01179, , https://doi.org/10.7554/eLife.01179.
[89] J.A. Denis, J. Nectoux, P.-J. Lamy, C. Rouillac Le Sciellour, H. Guermouche, A.-S. Alary, O. Kosmider, N. Sarafan-Vasseur, C. Jovelet, B. Busser, P. Nizard, V. Taly, F. Fina, Development of digital PCR molecular tests for clinical practice: principles, practical implementation and recommendations, Ann. Biol. Clin. (Paris). 76 (2018) 505–523, https://doi.org/10.1684/abc.2018.1372.
[90] Z. Tao, A. Shi, R. Li, Y. Wang, X. Wang, J. Zhao, Microarray bioinformatics in cancer- a review, J. BUON 22 (2019) 838–843 n.d (accessed June 27, 2019) http://www.ncbi.nlm.nih.gov/pubmed/29155508.
[91] S.A. Byron, K.R. Van Keuren-Jensen, D.M. Engelthaler, J.D. Carpten, D.W. Craig, Translating RNA sequencing into clinical diagnostics: opportunities and challenges, Nat. Rev. Genet. 17 (2016) 257–271, https://doi.org/10.1038/nrg.2016.10.
[92] C.S. Carlson, R.O. Emerson, A.M. Sherwood, C. Desmarais, M.-W. Chung, J.M. Parsons, M.S. Steen, M.A. LaMadrid-Herrmannsfeldt, D.W. Williamson, R.J. Livingston, D. Wu, B.L. Wood, M.J. Rieder, H. Robins, Using synthetic templates to design an unbiased multiplex PCR assay, Nat. Commun. 4 (2013) 2680, https://doi.org/10.1038/ncomms3680.
[93] R. Nault, K.A. Fader, T. Zacharewski, RNA-seq versus oligonucleotide array assessment of dose-dependent TCDD-elicited hepatic gene expression in mice, BMC Genomics. 16 (2015) 373, https://doi.org/10.1186/s12864-015-1527-z.
[94] A. Watson, A. Mazumder, M. Stewart, S. Balasubramanian, Technology for microarray analysis of gene expression, Curr. Opin. Biotechnol. 9 (1998) 609–614 (accessed June 27, 2019), http://www.ncbi.nlm.nih.gov/pubmed/9889134.
[95] D.D. Dalma-Weiszhausz, J. Warrington, E.Y. Tanimoto, C.G. Miyada, The affymetrix GeneChip® platform: an overview, Methods Enzymol (2006) 3–28, https://doi.org/10.1016/S0076-6879(06)10001-4.
[96] A. Teumer, C. Schurmann, A. Schillert, K. Schramm, A. Ziegler, Prokisch H. Analyzing illumina Gene expression microarray data obtained from human whole blood cell and blood monocyte samples, Methods Mol. Biol. (2016) 85–97, https://doi.org/10.1007/978-1-4939-3136-1_7.
[97] P.K. Wolber, P.J. Collins, A.B. Lucas, A. De Witte, K.W. Shannon, The agilent in situ-synthesized microarray platform, Methods Enzymol. (2006) 28–57, https://doi.org/10.1016/S0076-6879(06)10002-6.
[98] I. Petrov, M. Suntsova, O. Mutorova, M. Sorokin, A. Garazha, E. Ilnitskaya, P. Spirin, S. Larin, O. Kovalchuk, V. Prassolov, A. Zhavoronkov, A. Roumiantsev, A. Buzdin, Molecular pathway activation features of pediatric acute myeloid leukemia (AML) and acute lymphoblast leukemia (ALL) cells, Aging (Albany. NY). 8 (2016) 2936–2947, https://doi.org/10.18632/aging.101102.
[99] N. Borisov, M. Suntsova, M. Sorokin, A. Garazha, O. Kovalchuk, A. Aliper, E. Ilnitskaya, K. Lezhnina, M. Korzinkin, V. Tkachev, V. Saenko, Y. Saenko, D.G. Sokov, N.M. Gaifullin, K. Kashintsev, V. Shirokorad, I. Shabalina, A. Zhavoronkov, B. Mishra, C.R. Cantor, A. Buzdin, Data aggregation at the level of molecular pathways improves stability of experimental transcriptomic and proteomic data, Cell. Cycle. (2017), https://doi.org/10.1080/15384101.2017.1361068 0–0.
[100] N.F. Lahens, E. Ricciotti, O. Smirnova, E. Toorens, E.J. Kim, G. Baruzzo, K.E. Hayer, T. Ganguly, J. Schug, G.R. Grant, A comparison of illumina and ion torrent sequencing platforms in the context of differential gene expression, BMC Genomics. 18 (2017) 602, https://doi.org/10.1186/s12864-017-4011-0.
[101] N. Kono, K. Arakawa, Nanopore sequencing: review of potential applications in functional genomics, Dev. Growth Differ. (2019) 12608, https://doi.org/10.1111/dgd.12608 dgd..
[102] A.A. Buzdin, A.A. Zhavoronkov, M.B. Korzinkin, S.A. Roumiantsev, A.M. Aliper, L.S. Venkova, P.Y. Smirnov, N.M. Borisov, The OncoFinder algorithm for minimizing the errors introduced by the high-throughput methods of transcriptome analysis, Front. Mol. Biosci. 1 (2014) 8, https://doi.org/10.3389/fmolb.2014.00008.
[103] N. Borisov, I. Shabalina, V. Tkachev, M. Sorokin, A. Garazha, A. Pulin, I.I. Eremin, A. Buzdin, Shambhala: a platform-agnostic data harmonizer for gene expression data, BMC Bioinform.. 20 (2019) 66, https://doi.org/10.1186/s12859-019-2641-8.
[104] A. Sîrbu, G. Kerr, M. Crane, H.J. Ruskin, RNA-seq vs dual- and single-channel microarray data: sensitivity analysis for differential expression and clustering, PLoS One. 7 (2012) e50986, , https://doi.org/10.1371/journal.pone.0050986.
[105] W. Zhang, Y. Yu, F. Hertwig, J. Thierry-Mieg, W. Zhang, D. Thierry-Mieg, J. Wang, M. Fischer, et al., Comparison of RNA-seq and microarray-based models for clinical endpoint prediction, Genome Biol. 16 (2015) 133, https://doi.org/10.1186/s13059-015-0694-1.
[106] M.F. Rai, E.D. Tycksen, L.J. Sandell, R.H. Brophy, Advantages of RNA-seq compared to RNA microarrays for transcriptome profiling of anterior cruciate ligament tears, J. Orthop. Res. 36 (2017) 484–497, https://doi.org/10.1002/jor.23661.
[107] SEQC/MAQC-III Consortium, A Comprehensive Assessment of RNA-Seq Accuracy, Reproducibility and Information Content by the Sequencing Quality Control

Consortium 32 (2014), https://doi.org/10.1038/nbt.2957.
[108] M. Schena, D. Shalon, R.W. Davis, P.O. Brown, Quantitative monitoring of Gene expression patterns with a complementary DNA microarray, Science (80-.) 270 (1995) 467–470, https://doi.org/10.1126/science.270.5235.467.
[109] B. Lin, J. Wang, Y. Cheng, Recent patents and advances in the next-generation sequencing technologies, Recent. Pat. Biomed. Eng. 2008 (2008) 60–67 (accessed June 27, 2019), http://www.ncbi.nlm.nih.gov/pubmed/21709726.
[110] D. Castillo, J.M. Gálvez, L.J. Herrera, B.S. Román, F. Rojas, I. Rojas, Integration of RNA-seq data with heterogeneous microarray data for breast cancer profiling, BMC Bioinform.. 18 (2017) 506, https://doi.org/10.1186/s12859-017-1925-0.
[111] I. Petrov, M. Suntsova, E. Ilnitskaya, S. Roumiantsev, M. Sorokin, A. Garazha, P. Spirin, T. Lebedev, N. Gaifullin, S. Larin, O. Kovalchuk, D. Konovalov, V. Prassolov, A. Roumiantsev, A. Buzdin, Gene expression and molecular pathway activation signatures of MYCN-amplified neuroblastomas, Oncotarget. 8 (2017) 83768–83780, https://doi.org/10.18632/oncotarget.19662.
[112] D. Shepelin, M. Korzinkin, A. Vanyushina, A. Aliper, N. Borisov, R. Vasilov, N. Zhukov, D. Sokov, V. Prassolov, N. Gaifullin, A. Zhavoronkov, B. Bhullar, A. Buzdin, Molecular pathway activation features linked with transition from normal skin to primary and metastatic melanomas in human, Oncotarget. 1 (2016) 656–670, https://doi.org/10.18632/oncotarget.6394.
[113] S.K. Kilpinen, K.A. Ojala, O.P. Kallioniemi, Alignment of gene expression profiles from test samples against a reference database: New method for context-specific interpretation of microarray data, BioData Min. 4 (2011) 5, https://doi.org/10.1186/1756-0381-4-5.
[114] H. Dvinge, R.E. Ries, J.O. Ilagan, D.L. Stirewalt, S. Meshinchi, R.K. Bradley, Sample processing obscures cancer-specific alterations in leukemic transcriptomes, Proc. Natl. Acad. Sci. 111 (2014) 16802–16807, https://doi.org/10.1073/pnas.1413374111.
[115] J. Lonsdale, J. Thomas, M. Salvatore, R. Phillips, E. Lo, S. Shad, R. Hasz, H.F. Moore, et al., The genotype-tissue expression (GTEx) project, Nat. Genet. 45 (2013) 580–585, https://doi.org/10.1038/ng.2653.
[116] Cancer Genome Atlas Research Network JN, J.N. Weinstein, E.A. Collisson, G.B. Mills, K.R.M. Shaw, B.A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander, J.M. Stuart, The cancer genome atlas Pan-cancer analysis project, Nat. Genet. 45 (2013) 1113–1120, https://doi.org/10.1038/ng.2764.
[117] C.A. Davis, B.C. Hitz, C.A. Sloan, E.T. Chan, J.M. Davidson, I. Gabdank, J.A. Hilton, K. Jain, U.K. Baymuradov, A.K. Narayanan, K.C. Onate, K. Graham, S.R. Miyasato, T.R. Dreszer, J.S. Strattan, O. Jolanki, F.Y. Tanaka, J.M. Cherry, The encyclopedia of DNA elements (ENCODE): data portal update, Nucleic Acids Res. 46 (2018) D794–D801, https://doi.org/10.1093/nar/gkx1081.
[118] X. Huang, D.F. Stern, H. Zhao, Transcriptional profiles from paired Normal samples offer complementary information on cancer patient survival – evidence from TCGA Pan-cancer data, Sci. Rep. 6 (20567) (2016), https://doi.org/10.1038/srep20567.
[119] A.C. Jones, K.S. Antillon, S.M. Jenkins, S.N. Janos, H.N. Overton, D.S. Shoshan, E.G. Fischer, K.A. Trujillo, M. Bisoffi, Prostate Field cancerization: deregulated expression of macrophage inhibitory cytokine 1 (MIC-1) and platelet derived growth factor A (PDGF-A) in tumor adjacent tissue, PLoS One. 10 (2015) e0119314, , https://doi.org/10.1371/journal.pone.0119314.
[120] P. Casbas-Hernandez, X. Sun, E. Roman-Perez, M. D'Arcy, R. Sandhu, A. Hishida, K.K. McNaughton, X.R. Yang, L. Makowski, M.E. Sherman, J.D. Figueroa, M.A. Troester, Tumor intrinsic subtype Is reflected in cancer-adjacent tissue, Cancer Epidemiol. Biomarkers Prev. 24 (2015) 406–414, https://doi.org/10.1158/1055-9965.EPI-14-0934.
[121] Y. Zhao, P. Yu, R. Wu, Y. Ge, J. Wu, J. Zhu, R. Jia, Renal cell carcinoma-adjacent tissues enhance mobilization and recruitment of endothelial progenitor cells to promote the invasion of the neoplasm, Biomed. Pharmacother. 67 (2013) 643–649, https://doi.org/10.1016/j.biopha.2013.06.009.
[122] M. Suntsova, N. Gaifullin, D. Allina, A. Reshetun, X. Li, L. Mendeleeva, V. Surin, A. Sergeeva, P. Spirin, V. Prassolov, A. Morgan, A. Garazha, M. Sorokin, A. Buzdin, Atlas of RNA sequencing profiles for normal human tissues, Sci. Data. 6 (2019) 36, https://doi.org/10.1038/s41597-019-0043-4.
[123] S.-H. Lin, L. Beane, D. Chasse, K.W. Zhu, B. Mathey-Prevot, J.T. Chang, Cross-platform prediction of Gene expression signatures, PLoS One. 8 (2013) e79228, , https://doi.org/10.1371/journal.pone.0079228.
[124] S. Maouche, O. Poirier, T. Godefroy, R. Olaso, I. Gut, J.-P. Collet, G. Montalescot, F. Cambien, Performance comparison of two microarray platforms to assess differential gene expression in human monocyte and macrophage cells, BMC Genomics. 9 (2008) 302, https://doi.org/10.1186/1471-2164-9-302.
[125] Z. Wen, C. Wang, Q. Shi, Y. Huang, Z. Su, H. Hong, W. Tong, L. Shi, Evaluation of gene expression data generated from expired affymetrix GeneChip® microarrays using MAQC reference RNA samples, BMC Bioinform. 11 (2010) S10, https://doi.org/10.1186/1471-2105-11-S6-S10.
[126] L. Zhang, J. Zhang, G. Yang, D. Wu, L. Jiang, Z. Wen, M. Li, Investigating the concordance of Gene ontology terms reveals the intra- and inter-platform reproducibility of enrichment analysis, BMC Bioinform. 14 (2013) 143, https://doi.org/10.1186/1471-2105-14-143.
[127] H.F.M. Kamel, Al-Amodi HSAB, Exploitation of Gene expression and cancer biomarkers in paving the path to Era of personalized medicine, Genomics. Proteom. Bioinform. 15 (2017) 220–235, https://doi.org/10.1016/j.gpb.2016.11.005.
[128] A.M. Aliper, M.B. Korzinkin, N.B. Kuzmina, A.A. Zenin, L.S. Venkova, P.Y. Smirnov, A.A. Zhavoronkov, A.A. Buzdin, N.M. Borisov, Mathematical justification of expression-based pathway activation scoring (PAS), Methods Mol. Biol. 1613 (2017) 31–51, https://doi.org/10.1007/978-1-4939-7027-8_3.
[129] B.M. Bolstad, R.A. Irizarry, M. Astrand, T.P. Speed, A comparison of normalization methods for high density oligonucleotide array data based on variance and bias, Bioinformatics 19 (2003) 185–193 (accessed May 21, 2017), http://www.ncbi.nlm.nih.gov/pubmed/12538238.
[130] M.N. McCall, B.M. Bolstad, R.A. Irizarry, Frozen robust multiarray analysis (fRMA), Biostatistics. 11 (2010) 242–253, https://doi.org/10.1093/biostatistics/kxp059.
[131] S. Anders, W. Huber, Differential expression analysis for sequence count data, Genome Biol. 11 (2010) R106, https://doi.org/10.1186/gb-2010-11-10-r106.
[132] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550, https://doi.org/10.1186/s13059-014-0550-8.
[133] J. Rudy, F. Valafar, Empirical comparison of cross-platform normalization methods for gene expression data, BMC Bioinform. 12 (2011) 467, https://doi.org/10.1186/1471-2105-12-467.
[134] A.A. Shabalin, H. Tjelmeland, C. Fan, C.M. Perou, A.B. Nobel, Merging two gene-expression studies via cross-platform normalization, Bioinformatics 24 (2008) 1154–1160, https://doi.org/10.1093/bioinformatics/btn083.
[135] H. Huang, X. Lu, Y. Liu, P. Haaland, J.S. Marron, R/DWD: distance-weighted discrimination for classification, visualization and batch adjustment, Bioinformatics. 28 (2012) 1182–1183, https://doi.org/10.1093/bioinformatics/bts096.
[136] M. Benito, J. Parker, Q. Du, J. Wu, D. Xiang, C.M. Perou, J.S. Marron, Adjustment of systematic microarray data biases, Bioinformatics. 20 (2004) 105–114, https://doi.org/10.1093/bioinformatics/btg385.
[137] W.L. Walker, I.H. Liao, D.L. Gilbert, B. Wong, K.S. Pollard, C.E. McCulloch, L. Lit, F.R. Sharp, Empirical bayes accomodation of batch-effects in microarray data using identical replicate reference samples: application to RNA expression profiling of blood from duchenne muscular dystrophy patients, BMC Genomics. 9 (2008) 494, https://doi.org/10.1186/1471-2164-9-494.
[138] P. Warnat, R. Eils, B. Brors, Cross-platform analysis of cancer microarray data improves gene expression based classification of phenotypes, BMC Bioinform. 6 (2005) 265, https://doi.org/10.1186/1471-2105-6-265.
[139] H. Jiang, Y. Deng, H.-S. Chen, L. Tao, Q. Sha, J. Chen, C.-J. Tsai, S. Zhang, Joint analysis of two microarray gene-expression data sets to select lung adenocarcinoma marker genes, BMC Bioinform.. 5 (2004) 81, https://doi.org/10.1186/1471-2105-5-81.
[140] X.-Q. Xia, M. McClelland, S. Porwollik, W. Song, X. Cong, Y. Wang, WebArrayDB: cross-platform microarray data analysis and public data repository, Bioinformatics. 25 (2009) 2425–2429, https://doi.org/10.1093/bioinformatics/btp430.
[141] A.G. Deshwar, Q. Morris, PLIDA: cross-platform gene expression normalization using perturbed topic models, Bioinformatics. 30 (2014) 956–961, https://doi.org/10.1093/bioinformatics/btt574.
[142] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T.L. Madden, BLAST+: architecture and applications, BMC Bioinformatics. 10 (2009) 421, https://doi.org/10.1186/1471-2105-10-421.
[143] J.A. Scolnick, M. Dimon, I.-C. Wang, S.C. Huelga, D.A. Amorese, An efficient method for identifying Gene fusions by targeted RNA sequencing from fresh frozen and FFPE samples, PLoS One. 10 (2015) e0128916, , https://doi.org/10.1371/journal.pone.0128916.
[144] A. Esteve-Codina, O. Arpi, M. Martinez-García, E. Pineda, M. Mallo, M. Gut, C. Carrato, A. Rovira, R. Lopez, A. Tortosa, M. Dabad, S. Del Barco, S. Heath, S. Bagué, T. Ribalta, F. Alameda, N. de la Iglesia, C. Balaña, GLIOCAT Group, A comparison of RNA-seq results from paired formalin-fixed paraffin-embedded and fresh-frozen glioblastoma tissue samples, PLoS One. 12 (2017) e0170632, , https://doi.org/10.1371/journal.pone.0170632.
[145] A. Buzdin, M. Sorokin, E. Poddubskaya, N. Borisov, High-throughput mutation data now complement transcriptomic profiling: advances in molecular pathway activation analysis approach in cancer biology, Cancer Inf.. 18 (2019) 1176935119838844, , https://doi.org/10.1177/1176935119838844.
[146] V. Tkachev, M. Sorokin, A. Mescheryakov, A. Simonov, A. Garazha, A. Buzdin, I. Muchnik, N. Borisov, FLOating-window projective separator (FloWPS): A data trimming tool for support vector machines (SVM) to improve robustness of the classifier, Front. Genet. 9 (2019) 717, https://doi.org/10.3389/fgene.2018.00717.
[147] C.D. Savci-Heijink, H. Halfwerk, J. Koster, M.J. Van de Vijver, Association between gene expression profile of the primary tumor and chemotherapy response of metastatic breast cancer, BMC Cancer. 17 (2017) 755, https://doi.org/10.1186/s12885-017-3691-9.
[148] A. Naruke, M. Azuma, A. Takeuchi, K. Ishido, C. Katada, T. Sasaki, K. Higuchi, S. Tanabe, M. Saegusa, W. Koizumi, Comparison of site-specific gene expression levels in primary tumors and synchronous lymph node metastases in advanced gastric cancer, Gastric Cancer. 18 (2015) 262–270, https://doi.org/10.1007/s10120-014-0357-z.
[149] E. Eden, R. Navon, I. Steinfeld, D. Lipson, Z. Yakhini, GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists, BMC Bioinform.. 10 (2009) 48, https://doi.org/10.1186/1471-2105-10-48.
[150] D.W. Huang, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (2009) 44–57, https://doi.org/10.1038/nprot.2008.211.
[151] A. Buzdin, M. Sorokin, A. Garazha, M. Sekacheva, E. Kim, N. Zhukov, Y. Wang, X. Li, S. Kar, C. Hartmann, A. Samii, A. Giese, N. Borisov, Molecular pathway activation – New type of biomarkers for tumor morphology and personalized selection of target drugs, Semin. Cancer Biol. (2018), https://doi.org/10.1016/j.semcancer.2018.06.003.
[152] A.A. Buzdin, V. Prassolov, A.A. Zhavoronkov, N.M. Borisov, Bioinformatics meets biomedicine: OncoFinder, a quantitative approach for interrogating molecular pathways using Gene expression data, Methods Mol. Biol. 1613 (2017) 53–83,

https://doi.org/10.1007/978-1-4939-7027-8_4.
[153] A.V. Artcibasova, M.B. Korzinkin, M.I. Sorokin, P.V. Shegay, A.A. Zhavoronkov, N. Gaifullin, B.Y. Alekseev, N.V. Vorobyev, D.V. Kuzmin, AD. Kaprin, N.M. Borisov, A.A. Buzdin, MiRImpact, a new bioinformatic method using complete microRNA expression profiles to assess their overall influence on the activity of intracellular molecular pathways, Cell. Cycle. 15 (2016) 689–698, https://doi.org/10.1080/15384101.2016.1147633.
[154] C. Sammut, G.I. Webb (Eds.), Encyclopedia of Machine Learning, Springer US, Boston, MA, 2010, , https://doi.org/10.1007/978-0-387-30164-8.
[155] D. Chicco, Ten quick tips for machine learning in computational biology, BioData Min. 10 (2017) 35, https://doi.org/10.1186/s13040-017-0155-3.
[156] T. Fawcett, An introduction to ROC analysis, Pattern Recognit. Lett. 27 (2006) 861–874, https://doi.org/10.1016/J.PATREC.2005.10.010.
[157] P. Perruchet, R. Peereman, The exploitation of distributional information in syllable processing, J. Neurolinguistics. 17 (2004) 97–119, https://doi.org/10.1016/S0911-6044(03)00059-9.
[158] P. Bartlett, J. Shawe-taylor, Generalization performance of support vector machines and other pattern classifiers, Adv. Kernel Methods Support Vector Learn. (1999) 43–54 Doi 10.1.1.42.6950 (accessed June 26, 2019) http://citeseerx.ist.psu.edu/viewdoc/summary?.
[159] M. Minsky, S. Papert, Perceptrons: An Introduction to Computational Geometry, MIT Press, 1988.
[160] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
[161] T.M. Mitchell, Machine Learning, New York (1997).
[162] X. Robin, N. Turck, A. Hainard, F. Lisacek, J.-C. Sanchez, M. Müller, Bioinformatics for protein biomarker panel classification: what is needed to bring biomarker panels into in vitro diagnostics? Expert rev, Proteomics. 6 (2009) 675–689, https://doi.org/10.1586/epr.09.83.
[163] H.-J. Cho, S. Lee, Y.G. Ji, D.H. Lee, Association of specific gene mutations derived from machine learning with survival in lung adenocarcinoma, PLoS One. 13 (2018) e0207204, , https://doi.org/10.1371/journal.pone.0207204.
[164] O. Soufan, D. Kleftogiannis, P. Kalnis, V.B. Bajic, DWFS: a wrapper feature selection tool based on a parallel genetic algorithm, PLoS One. 10 (2015) e0117988, , https://doi.org/10.1371/journal.pone.0117988.
[165] Z. Wang, H. Yang, Z. Wu, T. Wang, W. Li, Y. Tang, G. Liu, In silico prediction of blood-brain barrier permeability of compounds by machine learning and resampling methods, ChemMedChem. 13 (2018) 2189–2201, https://doi.org/10.1002/cmdc.201800533.
[166] T. Turki, Z. Wei, Learning approaches to improve prediction of drug sensitivity in breast cancer patients, in: 2016 38th, Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., IEEE (2016) 3314–3320, https://doi.org/10.1109/EMBC.2016.7591437.
[167] S.M. Ayyad, A.I. Saleh, L.M. Labib, Gene expression cancer classification using modified K-nearest neighbors technique, Biosystems. 176 (2019) 41–51, https://doi.org/10.1016/j.biosystems.2018.12.009.
[168] G. Mulligan, C. Mitsiades, B. Bryant, F. Zhan, W.J. Chng, S. Roels, E. Koenig, A. Fergus, Y. Huang, P. Richardson, W.L. Trepicchio, A. Broyl, P. Sonneveld, J.D. Shaughnessy, P.L. Bergsagel, D. Schenkein, D.-L. Esseltine, A. Boral, K.C. Anderson, Gene expression profiling and correlation with outcome in clinical trials of the proteasome inhibitor bortezomib, Blood. 109 (2007) 3177–3188, https://doi.org/10.1182/blood-2006-09-044974.
[169] N. Borisov, V. Tkachev, M. Suntsova, O. Kovalchuk, A. Zhavoronkov, I. Muchnik, A. Buzdin, A method of gene expression data transfer from cell lines to cancer patients for machine-learning prediction of drug efficiency, Cell. Cycle. 17 (2018) 486–491, https://doi.org/10.1080/15384101.2017.1417706.
[170] M.S. Brose, M.E. Cabanillas, E.E.W. Cohen, L.J. Wirth, T. Riehl, H. Yue, S.I. Sherman, E.J. Sherman, Vemurafenib in patients with BRAFV600E-positive metastatic or unresectable papillary thyroid cancer refractory to radioactive iodine: a non-randomised, multicentre, open-label, phase 2 trial, Lancet Oncol. 17 (2016) 1272–1282, https://doi.org/10.1016/S1470-2045(16)30166-8.
[171] A. Sgambato, F. Casaluce, P. Maione, C. Gridelli, Targeted therapies in non-small cell lung cancer: a focus on ALK/ROS1 tyrosine kinase inhibitors, Expert Rev. Anticancer. Ther. 18 (2018) 71–80, https://doi.org/10.1080/14737140.2018.1412260.
[172] R. Piskol, G. Ramaswami, J.B. Li, Reliable identification of genomic variants from RNA-seq data, Am. J. Hum. Genet. 93 (2013) 641–651, https://doi.org/10.1016/j.ajhg.2013.08.008.
[173] F. Koeppel, A. Bobard, C. Lefebvre, M. Pedrero, M. Deloger, Y. Boursin, C. Richon, R. Chen-Min-Tao, G. Robert, G. Meurice, E. Rouleau, S. Michiels, C. Massard, J.-Y. Scoazec, E. Solary, J.-C. Soria, F. André, L. Lacroix, Added value of whole-exome and transcriptome sequencing for clinical molecular screenings of advanced cancer patients with solid tumors, Cancer J. 24 (2018) 153–162, https://doi.org/10.1097/PPO.0000000000000322.
[174] R.M. Samstein, C.-H. Lee, A.N. Shoushtari, M.D. Hellmann, R. Shen, Y.Y. Janjigian, D.A. Barron, L.G.T. Morris, et al., Tumor mutational load predicts survival after immunotherapy across multiple cancer types, Nat. Genet. 51 (2019) 202–206, https://doi.org/10.1038/s41588-018-0312-8.
[175] R. Büttner, J.W. Longshore, F. López-Ríos, S. Merkelbach-Bruse, N. Normanno, E. Rouleau, F. Penault-Llorca, Implementing TMB measurement in clinical practice: considerations on assay requirements, ESMO Open. 4 (2019) e000442, , https://doi.org/10.1136/esmoopen-2018-000442.
[176] V. Lazar, E. Rubin, S. Depil, Y. Pawitan, J.-F. Martini, J. Gomez-Navarro, A. Yver, R. Kurzrock, et al., A simplified interventional mapping system (SIMS) for the selection of combinations of targeted treatments in non-small cell lung cancer, Oncotarget. 6 (2015) 14139–14152, https://doi.org/10.18632/oncotarget.3741.
[177] A. Artemov, A. Aliper, M. Korzinkin, K. Lezhnina, L. Jellen, N. Zhukov, S. Roumiantsev, N. Gaifullin, A. Zhavoronkov, N. Borisov, A. Buzdin, A method for predicting target drug efficiency in cancer based on the analysis of signaling pathway activation, Oncotarget. 6 (2015) 29347–29356, https://doi.org/10.18632/oncotarget.5119.
[178] S.-J. Kim, S. Kim, D.-W. Kim, M. Kim, B. Keam, T.M. Kim, Y. Lee, J. Koh, Y.K. Jeon, D.S. Heo, Alterations in PD-L1 expression associated with acquisition of resistance to ALK inhibitors in ALK-rearranged lung cancer, Cancer Res. Treat. (2018), https://doi.org/10.4143/crt.2018.486.
[179] N.R. Smith, C. Womack, A matrix approach to guide IHC-based tissue biomarker development in oncology drug discovery, J. Pathol. 232 (2014) 190–198, https://doi.org/10.1002/path.4262.
[180] J. Adam, N. Le Stang, I. Rouquette, A. Cazes, C. Badoual, H. Pinot-Roussel, L. Tixier, C. Danel, F. Damiola, D. Damotte, F. Penault-Llorca, S. Lantuéjoul, Multicenter harmonization study for PD-L1 IHC testing in non-small-cell lung cancer, Ann. Oncol. 29 (2018) 953–958, https://doi.org/10.1093/annonc/mdy014.
[181] J.M. Conroy, S. Pabla, M.K. Nesline, S.T. Glenn, A. Papanicolau-Sengos, B. Burgher, J. Andreas, V. Giamo, Y. Wang, F.L. Lenzo, W. Bshara, M. Khalil, G.K. Dy, K.G. Madden, K. Shirai, K. Dragnev, L.J. Tafe, J. Zhu, M. Labriola, D. Marin, S.J. McCall, J. Clarke, D.J. George, T. Zhang, M. Zibelman, P. Ghatalia, I. Araujo-Fernandez, L. de la Cruz-Merino, A. Singavi, B. George, A.C. MacKinnon, J. Thompson, R. Singh, R. Jacob, D. Kasuganti, N. Shah, R. Day, L. Galluzzi, M. Gardner, C. Morrison, Next generation sequencing of PD-L1 for predicting response to immune checkpoint inhibitors, J. Immunother. Cancer. 7 (2019) 18, https://doi.org/10.1186/s40425-018-0489-5.
[182] N.M. Borisov, N.V. Terekhanova, A.M. Aliper, L.S. Venkova, P.Y. Smirnov, S. Roumiantsev, M.B. Korzinkin, A.A. Zhavoronkov, A.A. Buzdin, Signaling pathways activation profiles make better markers of cancer than expression of individual genes, Oncotarget 5 (2014) 10198–10205 (accessed October 5, 2015), http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4259415&tool=pmcentrez&rendertype=abstract.
[183] L. Yang, D. Roberts, M. Takhar, N. Erho, B.A.S. Bibby, N. Thiruthaneeswaran, V. Bhandari, W.-C. Cheng, S. Haider, A.M.B. McCorry, D. McArt, S. Jain, M. Alshalalfa, A. Ross, E. Schaffer, Jeffrey Den RB, R. Karnes, E. Klein, P.J. Hoskin, S.J. Freedland, A.D. Lamb, D.E. Neal, F.M. Buffa, R.G. Bristow, P.C. Boutros, E. Davicioni, A. Choudhury, C.M.L. West, Development and validation of a 28-gene hypoxia-related prognostic signature for localized prostate cancer, EBioMedicine. 31 (2018) 182–189, https://doi.org/10.1016/j.ebiom.2018.04.019.
[184] D. Sun, J. Chen, L. Liu, G. Zhao, P. Dong, B. Wu, J. Wang, L. Dong, Establishment of a 12-gene expression signature to predict colon cancer prognosis, PeerJ. 6 (2018) e4942, https://doi.org/10.7717/peerj.4942.
[185] C.M. Perou, T. Sørlie, M.B. Eisen, M. van de Rijn, S.S. Jeffrey, C.A. Rees, J.R. Pollack, D.T. Ross, H. Johnsen, L.A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S.X. Zhu, P.E. Lønning, A.L. Børresen-Dale, P.O. Brown, D. Botstein, Molecular portraits of human breast tumours, Nature. 406 (2000) 747–752, https://doi.org/10.1038/35021093.
[186] J. Guinney, R. Dienstmann, X. Wang, A. de Reyniès, A. Schlicker, C. Soneson, L. Marisa, P. Roepman, G. Nyamundanda, P. Angelino, B.M. Bot, J.S. Morris, I.M. Simon, S. Gerster, E. Fessler, F. De Sousa E Melo, E. Missiaglia, H. Ramay, D. Barras, K. Homicsko, D. Maru, G.C. Manyam, B. Broom, V. Boige, B. Perez-Villamil, T. Laderas, R. Salazar, J.W. Gray, D. Hanahan, J. Tabernero, R. Bernards, S.H. Friend, P. Laurent-Puig, J.P. Medema, A. Sadanandam, L. Wessels, M. Delorenzi, S. Kopetz, L. Vermeulen, S. Tejpar, The consensus molecular subtypes of colorectal cancer, Nat. Med. 21 (2015) 1350–1356, https://doi.org/10.1038/nm.3967.
[187] J. Stebbing, H. Singh Wasan, Decoding metastatic colorectal cancer to improve clinical decision making, J. Clin. Oncol. (2019) JCO1901185, https://doi.org/10.1200/JCO.19.01185.
[188] M.J. van de Vijver, Y.D. He, L.J. van't Veer, H. Dai, A.A.M. Hart, D.W. Voskuil, G.J. Schreiber, J.L. Peterse, C. Roberts, M.J. Marton, M. Parrish, D. Atsma, A. Witteveen, A. Glas, L. Delahaye, T. van der Velde, H. Bartelink, S. Rodenhuis, E.T. Rutgers, S.H. Friend, R. Bernards, A gene-expression signature as a predictor of survival in breast cancer, N. Engl. J. Med. 347 (2002) 1999–2009, https://doi.org/10.1056/NEJMoa021967.
[189] F. Cardoso, L.J. van't Veer, J. Bogaerts, L. Slaets, G. Viale, S. Delaloge, J.-Y. Pierga, E. Brain, S. Causeret, M. DeLorenzi, A.M. Glas, V. Golfinopoulos, T. Goulioti, S. Knox, E. Matos, B. Meulemans, P.A. Neijenhuis, U. Nitz, R. Passalacqua, P. Ravdin, I.T. Rubio, M. Saghatchian, T.J. Smilde, C. Sotiriou, L. Stork, C. Straehle, G. Thomas, A.M. Thompson, J.M. van der Hoeven, P. Vuylsteke, R. Bernards, K. Tryfonidis, E. Rutgers, M. Piccart, MINDACT investigators. 70-Gene signature as an aid to treatment decisions in early-stage breast cancer, N. Engl. J. Med. 375 (2016) 717–729, https://doi.org/10.1056/NEJMoa1602253.
[190] A. Nicolini, P. Ferrari, M.J. Duffy, Prognostic and predictive biomarkers in breast cancer: past, present and future, Semin. Cancer Biol. 52 (2018) 56–73, https://doi.org/10.1016/j.semcancer.2017.08.010.
[191] S. Paik, S. Shak, G. Tang, C. Kim, J. Baker, M. Cronin, F.L. Baehner, M.G. Walker, D. Watson, T. Park, W. Hiller, E.R. Fisher, D.L. Wickerham, J. Bryant, N. Wolmark, A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer, N. Engl. J. Med. 351 (2004) 2817–2826, https://doi.org/10.1056/NEJMoa041588.
[192] J.A. Sparano, R.J. Gray, D.F. Makower, K.I. Pritchard, K.S. Albain, D.F. Hayes, C.E. Geyer, E.C. Dees, M.P. Goetz, J.A. Olson, T. Lively, S.S. Badve, T.J. Saphner, L.I. Wagner, T.J. Whelan, M.J. Ellis, S. Paik, W.C. Wood, P.M. Ravdin, M.M. Keane, H.L. Gomez Moreno, P.S. Reddy, T.F. Goggins, I.A. Mayer, A.M. Brufsky, D.L. Toppmeyer, V.G. Kaklamani, J.L. Berenberg, J. Abrams, G.W. Sledge, Adjuvant chemotherapy guided by a 21-Gene expression assay in breast cancer, N.

Engl. J. Med. 379 (2018) 111–121, https://doi.org/10.1056/NEJMoa1804710.
[193] T.I. Rogounovitch, A. Bychkov, M. Takahashi, N. Mitsutake, M. Nakashima, A.V. Nikitski, T. Hayashi, M. Hirokawa, K. Ishigaki, K. Shigematsu, T. Bogdanova, M. Matsuse, E. Nishihara, S. Minami, K. Yamanouchi, M. Ito, T. Kawaguchi, H. Kondo, N. Takamura, Y. Ito, A. Miyauchi, F. Matsuda, S. Yamashita, V.A. Saenko, The common genetic variant rs944289 on chromosome 14q13.3 associates with risk of both malignant and benign thyroid tumors in the Japanese population, Thyroid 25 (2015) 333–340, https://doi.org/10.1089/thy.2014.0431.
[194] M. Nishino, M. Nikiforova, Update on molecular testing for cytologically indeterminate thyroid nodules, Arch. Pathol. Lab. Med. 142 (2018) 446–457, https://doi.org/10.5858/arpa.2017-0174-RA.
[195] R. Samsonov, V. Burdakov, T. Shtam, Z. Radzhabova, D. Vasilyev, E. Tsyrlina, S. Titov, M. Ivanov, L. Berstein, M. Filatov, N. Kolesnikov, H. Gil-Henn, A. Malek, Plasma exosomal miR-21 and miR-181a differentiates follicular from papillary thyroid cancer, Tumour Biol. 37 (2016) 12011–12021, https://doi.org/10.1007/s13277-016-5065-3.
[196] S.E. Titov, P.S. Demenkov, M.K. Ivanov, E.S. Malakhina, T.L. Poloz, E.V. Tsivlikova, M.S. Ganzha, S.P. Shevchenko, L.F. Gulyaeva, N.N. Kolesnikov, Selection and validation of miRNAs as normalizers for profiling expression of microRNAs isolated from thyroid fine needle aspiration smears, Oncol. Rep. 36 (2016) 2501–2510, https://doi.org/10.3892/or.2016.5113.
[197] J. Rodon, J.-C. Soria, R. Berger, W.H. Miller, E. Rubin, A. Kugel, A. Tsimberidou, P. Saintigny, A. Ackerstein, I. Braña, Y. Loriot, M. Afshar, V. Miller, F. Wunder, C. Bresson, J.-F. Martini, J. Raynaud, J. Mendelsohn, G. Batist, A. Onn, J. Tabernero, R.L. Schilsky, V. Lazar, J.J. Lee, R. Kurzrock, Genomic and transcriptomic profiling expands precision cancer medicine: the WINTHER trial, Nat. Med. 25 (2019) 751–758, https://doi.org/10.1038/s41591-019-0424-4.
[198] V. Tkachev, M. Sorokin, A. Garazha, N.B.A. Borisov, Oncobox method for scoring efficiencies of anticancer drugs based on gene expression data, Methods Mol. Biol. (2019) In press.
[199] A. Buzdin, M. Sorokin, A. Glusker, A. Garazha, E. Poddubskaya, V. Shirokorad, D. Naskhletashvili, K. Kashintsev, D. Sokov, M. Suntsova, S. Roumiantsev, O. Kovalchuk, N. Gaifullin, V. Tkachev, N. Borisov, Activation of intracellular signaling pathways as a new type of biomarkers for selection of target anticancer drugs, J. Clin. Oncol. 35 (2017), https://doi.org/10.1200/JCO.2017.35.15_suppl.e23142 e23142–e23142.
[200] E. Poddubskaya, A. Buzdin, A. Garazha, M. Sorokin, A. Glusker, A. Aleshin, D. Allina, A. Moiseev, M. Sekacheva, M. Suntsova, E.L. Kim, N. Gaifullin, V. Tkachev, A. Giese, V. Barbara, P. Borger, N. Borisov, Oncobox, gene expression-based second opinion system for predicting response to treatment in advanced solid tumors, J. Clin. Oncol. 37 (2019), https://doi.org/10.1200/JCO.2019.37.15_suppl.e13143 e13143–e13143.
[201] E.L. Kim, A. Buzdin, M. Sorokin, E. Poddubskaya, A. Poddubskiy, S.R. Kantelhardt, D. Kalasauskas, B. Sprang, S.E. Bikar, A. Giese, RNA-sequencing and bioinformatic analysis to pre-assess sensitivity to targeted therapeutics in recurrent glioblastoma, J. Clin. Oncol. 37 (2019), https://doi.org/10.1200/JCO.2019.37.15_suppl.e13533 e13533–e13533.
[202] E.V. Poddubskaya, M.P. Baranova, D.O. Allina, M.I. Sekacheva, L.A. Makovskaia, D.E. Kamashev, M.V. Suntsova, V.S. Barbara, I.N. Kochergina-Nikitskaya, A.A. Aleshin, Personalized prescription of imatinib in recurrent granulosa cell tumor of the ovary: case report, Cold Spring Harb. Mol. Case Stud. 5 (2019) a003434, , https://doi.org/10.1101/mcs.a003434.
[203] E.V. Poddubskaya, M.P. Baranova, D.O. Allina, P.Y. Smirnov, E.A. Albert, A.P. Kirilchev, A.A. Aleshin, M.I. Sekacheva, M.V. Suntsova, Personalized prescription of tyrosine kinase inhibitors in unresectable metastatic cholangiocarcinoma, Exp. Hematol. Oncol. 7 (2018) 21, https://doi.org/10.1186/s40164-018-0113-x.
[204] P. Spirin, T. Lebedev, N. Orlova, A. Morozov, N. Poymenova, S.E. Dmitriev, A. Buzdin, C. Stocking, O. Kovalchuk, V. Prassolov, Synergistic suppression of t(8;21)-positive leukemia cell growth by combining oridonin and MAPK1/ERK2 inhibitors, Oncotarget. 8 (2017) 56991–57002, https://doi.org/10.18632/oncotarget.18503.
[205] P.V. Spirin, T.D. Lebedev, N.N. Orlova, S. Gornostaeva a, M.M. Prokofijeva, Nikitenko N a, S.E. Dmitriev, Buzdin a a, N.M. Borisov, M. Aliper a, V. Garazha a, P.M. Rubtsov, C. Stocking, V.S. Prassolov, Silencing AML1-ETO gene expression leads to simultaneous activation of both pro-apoptotic and proliferation signaling, Leukemia 28 (2014) 1–23, https://doi.org/10.1038/leu.2014.130.
[206] V. Comunanza, D. Corà, F. Orso, F.M. Consonni, E. Middonti, F. Di Nicolantonio, A. Buzdin, A. Sica, E. Medico, D. Sangiolo, D. Taverna, F. Bussolino, VEGF blockade enhances the antitumor effect of BRAF $^{V\ 600E}$ inhibition, EMBO Mol. Med. 9 (2017) 219–237, https://doi.org/10.15252/emmm.201505774.
[207] M. Sorokin, R. Kholodenko, M. Suntsova, G. Malakhova, A. Garazha, I. Kholodenko, E. Poddubskaya, D. Lantsov, I. Stilidi, P. Arhiri, A. Osipov, A. Buzdin, Oncobox bioinformatical platform for selecting potentially effective combinations of target cancer drugs using high-throughput gene expression data, Cancers (Basel). 10 (2018) 365, https://doi.org/10.3390/cancers10100365.
[208] M.A. Zolotovskaia, M.I. Sorokin, S.A. Roumiantsev, N.M. Borisov, A.A. Buzdin, Pathway instability is an effective new mutation-based type of cancer biomarkers, Front. Oncol. 8 (2018) 658, https://doi.org/10.3389/fonc.2018.00658.
[209] M.A. Zolotovskaia, M.I. Sorokin, A.A. Emelianova, N.M. Borisov, D.V. Kuzmin, P. Borger, A.V. Garazha, A.A. Buzdin, Pathway based analysis of mutation data is efficient for scoring target cancer drugs, Front. Pharmacol. 10 (2019) 1, https://doi.org/10.3389/fphar.2019.00001.