Practical aspects of NGS-based pathways analysis for personalized cancer science and medicine

Ekaterina A. Kotelnikova[1,2,3,*], Mikhail Pyatnitskiy[1,6,7,*], Anna Paleeva[1], Olga Kremenetskaya[1,5] and Dmitriy Vinogradov[1,2,4]

[1] Personal Biomedicine, Moscow, Russia

[2] A. A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia

[3] Institute Biomedical Research August Pi Sunyer (IDIBAPS), Hospital Clinic of Barcelona, Barcelona, Spain

[4] Lomonosov Moscow State University, Leninskie Gory, Moscow, Russia

[5] Center for Theoretical Problems of Physicochemical Pharmacology RAS, Moscow, Russia

[6] Orekhovich Institute of Biomedical Chemistry, Pogodinskaya, Moscow, Russia

[7] Pirogov Russian National Research Medical University, Ostrovityanova, Moscow, Russia

[*] This authors have contributed equally to this work

Correspondence to: Ekaterina A. Kotelnikova, *email*: ekotelnikova@gmail.com

Keywords: next generation sequencing (NGS), systems biology, precision oncology, personalized medicine, pathways

Received: October 02, 2015   *Accepted*: April 18, 2016   *Published*: May 14, 2016

ABSTRACT

Nowadays, the personalized approach to health care and cancer care in particular is becoming more and more popular and is taking an important place in the translational medicine paradigm. In some cases, detection of the patient-specific individual mutations that point to a targeted therapy has already become a routine practice for clinical oncologists. Wider panels of genetic markers are also on the market which cover a greater number of possible oncogenes including those with lower reliability of resulting medical conclusions. In light of the large availability of high-throughput technologies, it is very tempting to use complete patient-specific New Generation Sequencing (NGS) or other "omics" data for cancer treatment guidance. However, there are still no gold standard methods and protocols to evaluate them. Here we will discuss the clinical utility of each of the data types and describe a systems biology approach adapted for single patient measurements. We will try to summarize the current state of the field focusing on the clinically relevant case-studies and practical aspects of data processing.

INTRODUCTION

The concept of molecular mechanisms, affected differently in different cancer patients, has been considered as a key to the correct personalized cancer treatment choice [1, 2]. The common way to assess these differences is to take into account markers of certain pathway activity and therapy response. Around 50 of tumor-specific predictive labels have been approved for the companion diagnostics of cancer treatment [3, 4]. In some cases, strong hereditary genetic markers (i.e germline mutations in BRCA1/2) could lead to preventive surgical tissue resection, resulting in a reduced risk of tumor appearance [5, 6].

In addition to the relatively simple single-gene tests widely used by practical oncologists, targeted sequencing panels that include several tens or hundreds of cancer-related genes are also available on the market. They could be used for clinical decisions and are offered by companies like Caris, Foundation Medicine, Personal Genome Diagnostics and others [7, 8] (Table 1). These extended tests mainly rely on the NGS technologies [9, 10] with high coverage of genomic regions of interest and can better describe molecular changes potentially leading to disturbance of cancer pathways.

Increasing availability of genetic testing facilities has led to the so-called basket trial studies, where selection of patient-specific cancer treatment is defined by tumor

Table 1: Companies, technologies and kits for precision oncology

Name	Website	Disease/Tissue specificity	Scope of coverage and methods	Description (detection with or without therapeutic interpretation)
Foundation Medicine	www.foundationmedicine.com/	Universal cancer panels: Solid tumors (Foundation one) and hematologic tumors (FoundationOne Heme)	Panels	Analysis of solid and Hematologic tumors– detection and interpretation of all class of genomic alterations (including base substitutions, InDels, CNAs, rearrangements and fusion genes)
Personal Genome Diagnostics (PGDX)	www.personalgenome.com/	Universal cancer Panels and tissue specific panel for NSCLC (LungSelect)	Full exome + panel (120 cancer genes)	Detection and interpretation of SNVs,InDels,CNAs and rearrangements
Ambry Genetics	www.ambrygen.com/	Universal cancer panels (solid tumors) + tissue specific cancer panels (ColoNext; OvaNext; PancNext; PGLNext; RenalNext)	Kits and cancer panels + exome and (mtDNA) genome	Detection and interpretation of gene SNVs, InDels, CNAs large rearrangements for specific types of cancer.
GeneDx	www.genedx.com	Universal cancer panels + tissue specific cancer panels (for breast, ovarian, colorectal, pancreatic, endometria cancers and Familial Cutaneous Malignant Melanoma)	Full exome (WES, NGS) + panels	Detection of SNVs, InDels. Deletion testing of mtDNA, detection of mtNDA SNVs.
NeoGenomics Laboratories	www.neogenomics.com/	Tissue specific cancer panels (for NSCLC; Melanoma; Colorectal Cancer)	Panels + IHC, FISH, Flow Cytometry, RT-PCR	Detection of SNVs, InDels, CNAs, rearrangements, fusions
Caris	http://www.carislifesciences.eu/solid-tumours	Universal cancer panel	Panels + IHC; CISH; FISH; RT-PCR; Sanger Sequencing, Pyro Sequencing; Fragment Analysis	Detection and interpretation of SNVs, CNAs, InDels, fusions and level of expression of protein biomarkers in solid tumors for therapeutic decision support and clinical trials matching.
Myriad Genetics	https://www.myriad.com/	Tissue specific cancer panels (breast and ovarian cancer)	Panel for BRCA1, BRCA2	Detection of gene mutations
Quest Diagnostics	www.questdiagnostics.com/	Universal cancer panel	Panel	Detection of SNVs and InDels;
GPS@WUSTL	http://gps.wustl.edu/	Universal cancer panels	Panel	Detection of SNVs and InDels
Arup Laboratories	https://www.aruplab.com	Tissue specific cancer panels for gastrointestinal cancer	Panel + IHC, FISH, and PCR	Screening, risk prediction, diagnosis, prognosis, monitoring, pharmacogenomics, and therapeutic triage of malignancies. Detection of SNVs, InDels, chromosomal alterations and level of expression of oncomarkers.

MolecularHealth	www. molecularhealth. com/	Universal cancer panels	Exome+panel (over 500 cancer-related genes), Comparative Genomic Hybridization (aCGH) used as an additional test	Interpretation of whole exome analysis data, detection and interpretation of gene alterations. Integration and interpretation of biological, medical and drug Response information.
Personalis	www.personalis. com/	Universal cancer panels (solid tumors)	Panels (more than 1,300 cancer genes and more than 200 miRNA genes)+ exome (WES)+ transcriptome	Detection of SNVs, InDels, CNAs, fusion genes, LOH, gene expression profiling, low-level variant expression.
OncoDNA	www.oncodna.com/	Universal cancer panels (solid tumors)	Panels (OncoDEEP DX - 65 genes, with wide coverage of the KRAS, BRAF, EGFR; OncoDEEP Clinical - more than 400 genes; Plus Package - multi-platform approach to complete the characterization of the tumor, including FISH, PCR, ICH)	Detection and interpretation of SNVs, InDels, CNAs, translocations, microsatellite instability, DNA methylation, presence and activation of specific proteins. Integration of all the data, analysis of molecular networks, findings of the latest publications and generation of a comprehensive and intuitive report.
GenomOncology	www. genomoncology. com/	Universal cancer panels (solid tumors)	Bioinformatic service	Interpretation of NGS data (SNVs, InDels, CNAs, translocations and other structural variants) and translate the specific molecular profile of each patient's tumor genome into an actionable clinical report.
MI-ONCOSEQ Study (Michigan Oncology Sequencing Center,University of Michigan)	http://mctp. med.umich.edu/ physicians/mi-oncoseq-study	Universal cancer panels (solid tumors)	WES + transcriptome sequencing	Detection and interpretation of tumor somatic and germline SNVs, InDels, CNAs, gene fusions and rearrangements, gene expression alterations.
Genewiz	www.genewiz.com/	Cancer (solid tumors)	Cancer panels (OncoGxOne™+ Hot spot cancer panels), exome sequencing, whole genome sequencing, transcriptome (RNA-Seq)	Detection and interpretation of SNVs, InDels, CNAs, rearrangements, low-frequency aberrations, gene fusions, transcriptome analysis, identification of splice variants.
Neogenomics Laboratories	www.neogenomics. com/	Universal cancer panels (solid tumors)	Panels	Detection and interpretation of genomic alterations including SNVs, InDels, CNAs.
Emory Genetics Laboratory	http://geneticslab. emory.edu/index. html	Universal cancer panels (solid tumors)	WES + Panels	Detection and interpretation of Exome data: SNVs, InDels.
Paradigm Cancer Diagnostic (PCDx)	http://www. paradigmdx.org/	Universal cancer panels (solid tumors)	Exome, transcriptome (over 500 cancer-related genes)	Detection and interpretation of a patient tumor SNVs, CNAs, InDels, rearrangements and fusions, mRNA expression and protein expression.
Rosetta Genomics	www. rosettagenomics. com/	Universal cancer panels (solid tumors)	Transcriptome	Detection only. microRNA-based diagnostics service.
ThermoFisher	https://www. thermofisher.com/	Universal cancer panels (solid tumors)	Exome	Detection of SNVs, InDels, CNAs and gene fusions.
Swift Bioscience	www.swiftbiosci. com/	Universal cancer panels (solid tumors)	Kits for Illumina NGS and Ion Torrent Platforms; TP53 panel for Illumina Platform	Detection of genes aberrations: SNVs and methylation status, from variety of clinical sample types.

Illumina	http://www.illumina.com/content/dam/illumina-marketing/documents/products/other/cancer-panel-comparison-tool.pdf	Universal cancer panels (solid tumors)	Panels	Detection of germline or somatic SNVs in solid and myeloid tumors.
Asuragene	http://asuragen.com/	Universal cancer panel	Pan cancer kit QUANTIDEX™	Detection of the scope of variants reported by the panel including >1,600 known COSMIC variants, SNVs, InDels, and structural rearrangements targeted by the panel.
RainDance Technologies	http://raindancetech.com/	Universal cancer panel + Tissue specific panels (for acute myeloid leukemia (AML), myelodysplastic syndromes (MDS), myeloproliferative Neoplasms (MPN), myeloma.	Panels: ThunderBolts™ Cancer Panel (Interrogate mutations/hotspots in 50 oncogenes, tumor suppressors and drug resistance markers); ThunderBolts™Myeloid Panel (Target mutations/hotspots in 49 genes implicated in AML, MDS, MPN and myeloma diseases, including challenging genes such as CEBPA and NOTCH1.	Detection of SNVs in cancer related genes.

molecular profile, but not by its tissue of origin [11-14]. However, some researchers suggest that one should not consider only tumor gene alterations but also take into account its tissue specific features. [15-17] (Table 1).

The most comprehensive NGS-based cancer studies would implicate the screening of germline genome [11], tumor genome [11, 12], transcriptome [13] and methylome [14] in the search of potential cancer driving alterations [15]. Genome screening can be performed by capturing cancer-related genes only (NGS-based cancer panels), all transcribed regions (whole-exome sequencing, WES) or whole-genome sequencing (WGS) [16]. Despite the tempting idea of WGS usage and the decreasing cost of NGS technologies, large-scale WGS studies are still unaffordable for many research laboratories and clinical settings. Thus, targeted cancer panels with high coverage of the selected genes or WES with restricted coverage of all genes, depending on the tasks, could be the preferable choice of DNA analysis.

In 2013, the U.S. Food and Drug Administration (FDA) approved the use of the Illumina MiSeqDx platform for high throughput NGS in clinics [17]. This decision by the FDA has paved the way for future clinical diagnostic and prognostic use of NGS and the emergence of the Precision Oncology 3.0 paradigm [18]. Precision Oncology 3.0 encourages the usage of systems biology, including pan-omics data and reverse engineering methods for hypothesizing the putative molecular networks that drive a given patient's tumor and for the selection of cancer or non-cancer off-label therapies that are potentially beneficial in the studied case. This approach, despite the presence of dosage, toxicity, efficacy and ethical issues, could be a promising strategy for oncologists to choose between the available therapies or to provide an alternative treatment regimen to the patients unresponsive to the standard care in order to improve therapeutic response and to minimize adverse events [19].

At the moment there are more than 50 web-sites that suggest different approaches to personalized cancer care [20]. Most of these cancer care organizations are using NGS-based targeted sequencing panels (Table 1) for studies of cancer driving SNVs (Single Nucleotide Variants) and InDels (Insertions and Deletions), while Fluorescence in situ hybridization (FISH) [21-23] and immunohistochemistry [24](IHC) are the standard methods of choice for the detection of cancer-related translocations and specific expression markers. Despite the fact that it is also possible to identify all the above-mentioned clinically-relevant molecular events using NGS methods (Table 2), the gold standards of NGS data processing for cancer samples are still under development [25]. There remain a lot of problems to be solved, the main concept to be proved and the confirmed designs of the Precision Oncology 3.0 clinical trials to be defined.

Table 2: Main clinically relevant cancer events, detectable by NGS

Event type	Sample type	Tissue type	References
Germline mutations (SNV/InDel)	DNA, RNA	Control tissue or blood	[192–195]
Somatic mutations (SNV/InDel)	DNA, RNA	Tumor and control tissue (or blood)	[61,196–199]
Somatic copy number alterations (CNA)	DNA	Tumor and control tissue (or blood)	[200–202]
Gene fusions and other somatic structural variations (SV)	DNA, RNA	Tumor and control tissue (or blood)	[23,203,204]
Methylation pattern changes	DNA	Tumor and control tissue	[14,77,205]
Differential gene expression	RNA	Tumor and control tissue	[73,206]
Differential alternative splicing	RNA	Tumor and control tissue	[73,207]

ENCOURAGING CASE STUDIES AND DESIGN OF CLINICAL TRIALS

The increasing attention to the field of precision oncology is supported by encouraging NGS-based personalized treatment guidance case studies. Among the first was a case of 78-year-old male patient, diagnosed with adenocarcinoma of the tongue published in 2010 by Jones et al. [26]. The patient went through erlotinib treatment without a positive effect and had to go through further therapy. Analysis of omics data (genome and transcriptome sequencing from tumor and normal tissues), using Ingenuity Pathway Analysis software [27], KEGG pathways [28] and the DrugBank [29]drug target database, revealed gene expression changes relevant to the signaling pathways involved in cancerogenesis. Two potential driver genes, up-regulated RET and down-regulated PTEN, that were probably related to the ineffectiveness of erlotinib were found. Once the therapy was changed to sunitinib, the volume of the tumor started to decrease. However, after 5 months the tumor started progressing again and the patient was transferred to sorafenib and sulindac as alternative drugs. Next, genome and transcriptome sequencing of tumor samples from metastasis was performed. Omics data analysis revealed nine *de novo* mutations not present in the controls nor in the tumor samples prior to the therapy. Further exploration suggested that resistance to sunitinib and sorafenib could be explained by the acquired upregulation of both MAPK/ERK and PI3K/AKT pathways. Eventually, this analysis of omics data led to the hypothesis that only a cocktail of targeted drugs would be able to reduce the proliferation of the tumor cells. The authors additionally speculated that as sequencing costs continue to decline, whole genome characterization will become a routine part of cancer pathology.

Welch JS et al. [30] in 2011 has described the use of WGS in "real-time" diagnosis and detection of an oncogenic fusion gene created by an insertional event. Within seven weeks, the authors had completed the process of library generation, massive parallel sequencing, analysis, and validation of a novel fusion that created a classic PML-RARA bcr3 variant. These findings altered the medical management of the patient, who then received all-trans retinoic acid instead of an allogeneic stem cell transplant.

One of the most inspiring examples was published not in a scientific article, but in the "New York Times" journal [31] in 2012. Oncologist himself, Dr. Lukas Wartman was diagnosed with the same type of tumor that he studied, Acute Lymphoblastic Leukemia (ALL). He was treated with chemotherapy and received necessary stem cell transplants. That put him back in remission, but in several years he relapsed again with only 4 or 5 percent chance of survival. Whole genome and whole transcriptome sequencing was then performed at the same institution where Dr. Wartman worked. The actionable modification (overexpression of FLT3) was found, and the drug (sunitinib or Sutent) approved for treating advanced kidney cancer, was administrated at his own risk. After the treatment, the patient went into full remission and, according to the Washington University School of Medicine in St.Louis web-page, returned to his work as an Assistant Professor in Oncology, at the time this review was written.

Other successful clinical examples of genetic analysis for personalized medicine were published in 2014 by Caris company[32]. Using the Caris Molecular Intelligence (CMI) platform - a combination of genome sequencing, FISH (fluorescence in situ hybridization method) and PCR the authors analyzed two patients. The first was a 63-year-old man with the progressive metastatic prostate cancer, which caused considerable

Table 3: Tools used for different modification types prediction.

Variation type	Single sample variant detection tools	Somatic variant detection tools	Difficulties	Clinical usefulness
SNV DNA	VarScan[193], GATK [66]	VarScan [193], Mutect [208], Strelka [209], GATK: UnifiedGenotyper [66]	High coverage is required for mutations with low allelic fraction. Reference bias.	Used by most of the approved genetically-based drug indications
SNV RNA	RVBoost [197]	UNCqeR [62], SNPiR [198], eSNV-detect [210]	Without DNA data can be confused with RNA-editing sites. Insufficient coverage for weakly expressed genes.	Provides extra layer of information whether mutated gene is expressed
InDel	Pindel [211][1], Dindel [212]	VarScan [193], Strelka [209], Scalpel [213][2]	Surrounding SNVs can prohibit correct read alignment.	Can greatly impact protein function by inducing a frameshift or deleting a domain
CNA	EWT [214], CNV-seq [215], FREEC [216]	Control-FreeC [200], BICseq [217], Excavator [201], ABSOLUTE [218]	Low boundary precision when used on WES data [67]	Help in driver genes detection [219], can be linked with outcome prognosis [220]
Fusions	TopHat-Fusion [221], SOAPfuse [222]	SOAPfuse [222], PRADA [223]	Validation is highly recommended. Can be confused with splicing aberrations.	Often linked very tightly with a specific disease, thus alleviating diagnosis
Differential methylation	N/A	DMRcate [224], comb-p [225], eDMR [226]	Experimental costs are rather high. Low coverage of all genomic CpG sites for some methods.	Can be used as biomarker for prognosis and therapy response prediction [227]
Differential expression	N/A	DESeq2 [74], Cufflinks [73], baySeq [75], limma [228]	Reliable prediction requires several replicates for both tissues. Control sample should be of the same origin as the tumor	May be used for diagnosis, prognosis, therapy response prediction and monitoring [229,230]
Differential splicing	N/A	DEXSeq [231], DSGSeq [232], MATS [233], Cufflinks [73], limma [228]	Requirements for replicates count are higher than for expression analysis. Rare splicing events detection needs high coverage.	May provide information for development of diagnostic tests, evaluating therapy efficacy [234], with potential application as prognostic and predictive markers [235]

There are modifications by third-party that allows search for somatic indels
[2] Method used for searching somatic InDels is not mentioned in original paper
N/A - not applied.

pain. The researchers identified decreased expression of thymidylate synthase (TS) in the tumor. Since low TS expression is known to be associated with tumor sensitivity to fluoropyrimidines and other folate analogs, the drug therapy - pemetrexed - had been prescribed. As a result, the size of the metastases was reduced and the tumor PSA marker, LDH, was normalized. The patient's condition returned to normal. The treatment has been tolerated exceptionally well and no further admissions to the hospital became necessary. The second patient was a 49-year-old woman diagnosed with stage IV ovarian cancer. Surgery confirmed metastatic disease and the patient began standard treatment with a combination of intravenous paclitaxel and carboplatin and intraperitoneal docetaxel/cisplatin. During the time of that treatment, the patient had a partial response. Biopsy material was sent for CMI testing to identify any additional treatment options. The CMI report indicated potential benefit from

the combination treatment of irinotecan and cetuximab based on the expression profile of the patient's tumor. This combination decreased the level of the patient's cancer antigen 125 to normal and allowed it to stay normal over the course of the first 8 months of treatment. Unfortunately, toxicity effects led to the discontinuation of the therapy. However, the demonstration of the efficacy irinotecan and cetuximab, which are rarely used in ovarian cancer treatment, is of significant importance as it justifies further exploration of treatments guided by tumor profiling instead of using the histological diagnosis of the tumor alone.

Among the most important up-to-date advances in current Precision Oncology, one could name the massive molecular-profiling-based clinical trial studies published by researchers from France [33, 34] and the USA [35-37] and multinational consortium WINTHER [38].

One of these studies, a whole-exome sequencing precision medicine trial that captures a diverse range of patients with advanced treatment-resistant cancer and prospective 7-25 months clinical follow-up, was published by Beltran H et al. [36]. More than 90% of the patients were shown to harbor actionable or biologically informative alterations, although treatment guided by this information was only present in 5% of the cases because of the lack of patient access to clinical trials and/or off-label use of drugs. Similarly, the feasibility study published by the French researchers [34] is the first proof-of-concept multicentric randomized clinical trial (SHIVA) comparing targeted therapy based on tumour molecular profiling *vs* conventional therapy in patients with refractory cancer. The druggable molecular abnormalities on the level of mutations, gene copy number alterations or IHC analyses were found for the 38 out of 100 enrolled patients with metastatic cancer who failed standard therapy.

Other modern approaches to clinical trials, extensively sponsored by National Institute of Cancer (NCI), include both "genotype to phenotype" and "phenotype to genotype" initiatives [39]. In particular, the molecular profiling-based assignment of cancer therapy is a goal of clinical trials NCI-MATCH [40] and NCI-MPACT [37, 41]. One of the examples of "phenotype to genotype" initiatives is "Exceptional Responders", a clinical trial inspired by previous case reports [42, 43]. It implies a retrospective analysis of tumor molecular features that may explain why patients responded particularly well to a particular treatment.

Multinational clinical trial WINTHER (five countries, six sites, coordinated by the Worldwide Innovative Networking Consortium) used genomic assays in making treatment decisions. This trial has been launched in order to assess the efficacy of therapy determined by matching of "genomic diagnosis" with targeted drugs [38].

These trials are the first ones to use a randomized design to examine whether assigning treatment based on genomic tumor screening can improve the rate and duration of response in patients with advanced solid tumors. Despite the fact that efficacy results for them are not available yet, the organization of the corresponding pipelines and clinical trial settings is extremely important for further advances and clinical validation of precision medicine approaches [15].

SYSTEMS BIOLOGY PIPELINE FOR THE CANCER-RELATED NGS DATA

Systems biology is a holistic rather than a reductionist approach to understanding and controlling biological complexity [44, 45]. Using systems biology, researchers obtain, integrate and analyze complex datasets from multiple experimental sources and molecular levels using interdisciplinary approaches. Being applied to the cancer research, the goal of systems biology is to decipher the impact of genetic and epigenetic aberrations in cancer cells onto their homeostasis, intercommunication and response to the possible treatments [46].

In systems biology, the scientific community is mainly focused on statistical approaches [47] and on trying to identify the features characteristic for the specific group of patients, and not on the concept of studying individual patient. As a consequence, some of the available tools in the field are not appliable for truly personalized studies. For example, pathways identified by gene expression profiling using group analyses differ considerably in comparison to those identified by personalized analyses [48]. However, the systems biology approach is specifically important for precision oncology [49], since each tumor is unique in terms of genetics, epigenetics and pathological rewiring of signaling pathways. Modeling of patient-specific molecular processes could help medical doctors identify the most effective treatment, minimize toxicity and avoid unnecessary trials and errors.

To start the NGS-based systems biology pipeline, it is necessary to obtain DNA and/or RNA samples from the studied tissues. The most widespread type of samples is the fresh-frozen paraffin embedded (FFPE) block, although blood sample or fresh surgical material could also be used upon availability and tasks. Thus, DNA or RNA, extracted from the blood [50, 51] as well as from FFPE [52, 53] or fresh tissue [54, 55] could be used for the detection of genetic alterations. Since transcriptomics studies are more prone to tissue-specificity, RNA from FFPE samples or fresh tissues of the same origin is necessary for reliable identification of gene expression changes.

Just after the sample preparation and detection by the best applicable sequencing techniques, many steps of data analysis have to be performed to obtain personalized clinically-relevant information. In general, these steps can be classified into 3 main categories: Sequencing & Bioinformatics, Functional Annotation & Pharmacogenomics and Systems Biology & Data integration. Generalized systems biology pipeline for
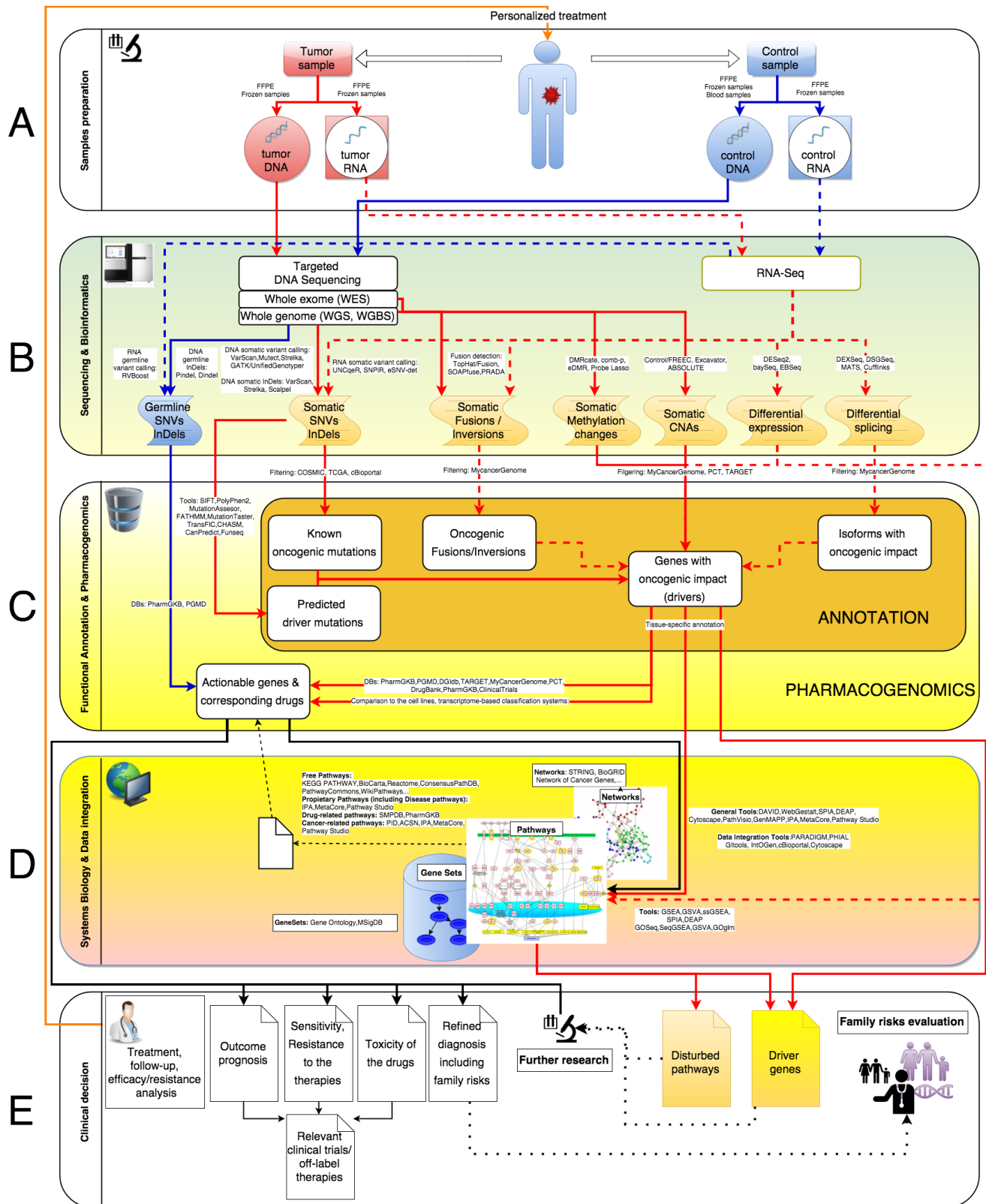
Figure 1: Generalized systems biology pipeline for the cancer-related NGS data processing. Here, solid blue and red lines correspond to DNA processing while dashed ones to RNA processing. Blue lines represent germline events, red ones - somatic. **A.** Sample preparation. Extraction of DNA and RNA from patient's tumor and normal tissue. **B.** Sequencing and Bioinformatics. Convert raw sequencing data into list of genetic variations. **C.** Functional annotation and Pharmacogenomics. **D.** Systems Biology and data integration. **E.** Clinical decision

the cancer-related NGS data processing is represented on Figure 1. Here we assume that all tasks necessary to identify potentially clinically relevant events (like SNVs calling, CNAs detection and so on, see Table 1), starting from read counts, could be addressed as bioinformatics tasks. The relevance of these events to cancer progression and response to the therapies could be inferred either by direct application of already available event-specific tools and databases (Functional Annotation & Pharmacogenomics), or by more sophisticated Systems Biology& Data integration approaches. Below we describe these categories in details.

SEQUENCING AND BIOINFORMATICS

Nowadays NGS is one of the most common high-throughput technologies due to its relatively low cost and high efficiency in processing of genetic material. The detailed description of NGS products and technologies is out of the scope of this paper but is thoroughly reviewed elsewhere [56, 57]. As with all high-throughput experimental methods, the data processing step is critical for obtaining correct results. Moreover, the impact of this step in the total cost of a sequencing project is steadily rising, while the share of experimental expenses is falling significantly [58]. Here we will cover main aspects of the NGS data processing, relevant to precision oncology.

The goal of the bioinformatic analysis can be expressed as 'to provide the exhaustive list of genetic features that are related to the occurrence of cancerous phenotype in the given sample'. This general task can be divided into several specific steps that are 1) NGS reads trimming and quality control; 2) mapping to a pre-built assembly; 3) alignment cleanup; 4) variant detection (or, broadly speaking, genome annotation). While the first three steps are rather technical (though important) and usually performed with small deviations from the generally common way, the last one is specific for each variation type. In Table 3 we summarize the information that can be useful during this stage (i.e. lists of the popular tools, noteworthy difficulties and some of the clinical implications), while a more detailed explanation is provided further.

The most frequent type of modifications is SNV (single nucleotide variations) [59]. Though many software packages have been developed for somatic SNV detection, the problem is far from being solved. The main source of difficulties is the low allelic fraction of mutations caused by tumor heterogeneity and polyclonality. Combined with practically reachable read coverage level this makes a large fraction of mutations indistinguishable from sequencing errors. Several recent papers note the high level of inconsistency among different tools predictions [60, 61]. To solve this issue, some authors propose simultaneous usage of several programs, which was proved to be advantageous [60]. Another problem is the

phenomenon called 'reference bias'- disposition of popular read mapping tools to discard or place incorrect reads with alternative alleles. Since reads with reference alleles are not affected, this leads to a decrease in maximum possible sensitivity, especially in weakly covered regions. One possible solution for this problem is to perform sequencing of RNA instead of DNA. The obvious benefits are higher coverage level for modestly expressed genes and potentially higher impact of all SNVs (mutations in non-expressed genes are less likely to be drivers). The main obstacles are the phenomenon of RNA-editing, which can lead to the appearance of false-positive calls in results, and low or zero coverage for weakly expressed alleles and regulatory regions. The best choice seems to be simultaneous sequencing of both DNA and RNA [62].

The second most common type of clinically useful events are short insertions and deletions (often referred as InDels). Modern read mappers often provide incorrect alignment in regions surrounding InDels, leading to a noticeable rise of error rate for both SNV and InDel calling [63, 64]. Thus realignment of reads in these regions is a crucial step of the bioinformatics analysis. Nevertheless it is not performed automatically by most popular read mappers because of its computational complexity. Some tools (e.g. ABRA [65] and HaplotypeCaller from GATK package [66]) incorporate another strategy - instead of mapping reads to reference genome, they perform local de-novo assembly.

While a single SNV or InDel act only on one gene, a CNA or a SV usually affects several of them. The process of detecting a set of CNA can be divided into three stages: estimating the copy-number for each locus, detecting true CNA boundaries by merging neighboring loci and then classifying the resulting CNAs. The first stage requires precise information about the local sequence properties in order to correct possible biases of the sequencing technology and the read mapping tool. Inferring proper boundaries is greatly impaired when WES strategy is used compared to WGS with accuracy being reduced up to two orders of magnitude [67]. For review of computational methods applicable to CNAs detection, see [68].

Gene fusions can often be heavily correlated with a specific cancer subtype (i.e. pathognomonic) or choice of targeted therapy. For example in a recent study [69] all patients with fibrolamellar hepatocellular carcinoma were found to have DNAJB1 and PRKACA genes fused, while no patients with other kinds of liver neoplasia had this modification. Additionally, gene fusions BCR-ABL and EML4-ALK are predictive markers to imatinib [70] and crizotinib [71] treatments, respectively. Therefore, fusion detection should be considered a crucial part of the diagnostic procedure. As SNVs and indels, gene fusions can be explored using either DNA or RNA data. RNA sequencing has several deficiencies (low expression levels of some fusions, inability to detect variations in regulatory regions), so the optimal strategy, again, seems to be the

simultaneous usage of DNA and RNA data [72].

Detection of differentially expressed genes between two tissue samples can be considered a quite mature area itself. Many methods were developed even before the advent of NGS technology in order to process expression data from hybridization microarrays, and while RNAseq expression data differs significantly in some aspects, the general idea mainly stays the same. Several of the most popular software packages include Cufflinks [73], DeSeq 2 [74] and baySeq [75]. And yet, some problems still persist. One of them is the very wide dynamic range of gene expression levels, which can yield noticeable bias in the results. Another source of issues is the presence of alternative splicing. Different transcripts of the same gene often perform quite distinct functions, which makes it important to separate gene isoforms during expression analysis. In its turn, detection of differences in alternative splicing events is complicated by the incomplete description of the splicing process even in healthy tissues.

Compared to expression analysis, evaluation of methylation data is rarely performed in cancer studies. Whereas the functional significance of DNA methylation for cancer has long been proved [76], genome-wide studies using comprehensive methods are still quite scarce [77], probably due to the high cost of experiments. In one of the recent examples, Stirzaker et al. suggests a possible connection of methylation patterns with outcome prognosis for triple-negative breast cancer [14].

Nowadays, next-generation sequencing (NGS) can also be used as a powerful tool for identification of rare events, e.g mosaicism. Some mutations acquired early in embryonic development that may be involved in cancer predisposition can be missed by less sensitive technologies [78]. The crucial point is the ability to detect low levels of mosaicism while accounting for the importance of tissue-specific mosaicism in disease and the potential increase of mosaicism frequency rate with age [79,80]. Mosaicism detection is important for individuals in the risk group or diagnosed with cancer. NGS based genetic testing may demonstrate levels of mosaicism much higher than the previously expectedfrequency. Mosaicsm may be observed in certain cases evenwithout apparent familial cancer history, as was demonstrated for gene APC and FAP (Familial adenomatous polyposis) patient, and BRCA in breast cancer patient. Today this approach is not widely applied in routine clinical practice and reports of somatic mosaicism detection are limited [81-83].

FUNCTIONAL ANNOTATION AND PHARMACOGENOMICS

A typical cancer sample contains several dozens of somatic mutations that may alter the functioning of the corresponding proteins. However, a relatively small fraction of genetic alterations leads to a small selective advantage of cancer cells and hence stimulates the tumor growth. Such alterations are called driver mutations [84], and their number is usually somewhere between two and eight per a tumor sample [19]. A subset of mutations may be "actionable", i.e. may have significant diagnostic, prognostic, or therapeutic implications in subsets of cancer patients [85]. On the contrary, the majority of the somatic mutations, so-called passenger mutations are a byproduct of the unstable cancer genome, and tend to not affect the fitness of tumor cells. Thus, they cannot serve as diagnostic and prognostic biomarkers [86]. However, there is some evidence that passenger mutations can be deleterious to cancer cells, altering the course of a tumor progression [87].

Since driver mutations provide growth advantage for the cancer cells, the most intuitive strategy to identify driver genes is to detect signals of evolutionary positive selection across tumor samples. Various approaches to quantify different evidences of selection pressure have been proposed. For example Tamborero et al. [88] employed several complementary methods including searching for genes with significant differences in mutation rates or enriched with mutations showing high functional impact, significant regional clustering or affecting phosphorylation-associated sites. This large-scale meta-analysis performed across 3,205 tumors produced list of 291 high-confidence driver genes. A similar type of analysis across 21 tumor types was done by Lawrence et al. [89] and integrated three independent signal types including enrichment of mutations in evolutionarily conserved sites. A total of 254 genes were identified.

Positively selected driver mutations are more likely to recur across multiple patients and tumor types [16]. Hence the first level of the event annotation may consist of filtering found somatic alterations and the identification of the previously reported ones. Recent advances in NGS methods have led to accumulation of thousands of publicly available cancer genomes. There are several sources of the relevant information, including the Catalogue of Somatic Mutations in Cancer, COSMIC [90] and the Cancer Genome Atlas, TCGA [91]. These huge amounts of data can be easily summarized with the help of other resources, such as cBioportal [92] or UCSC Cancer Genomics Browser [93]. However, frequency-based analysis has certain limitations in detecting driver mutations. Although several well-established cancer genes are mutated in a high proportion of tumours (like TP53, KRAS, BRAF, PTEN), most genes are mutated at intermediate and low frequencies (2-20%) [89].

It is important to emphasize the difference between predicting driver genes and individual driver mutations. Not all the alterations in the cancer-associated driver gene can be treated as driver mutations. An alteration in a proto-oncogene can be considered a driver mutation only if it leads to gene activation or results in a new function. Similarly, to claim a mutation is a driver, it should clearly impair functioning as a tumor suppressor. Hence, many

driver mutations have too low occurrence levels to be detected only by frequency-based analyses using currently available data. It will therefore be necessary to employ algorithms for driver/passenger prediction which consider the mutations' local functional and genomic context.

The challenge is to differentiate between driver and passenger mutations and rank the former according to their likelihood of promoting tumor progression. It is important to distinguish between individual alteration driver/passenger discrimination and prediction of the mutation impact upon a protein function. Many computational tools have been developed for the latter problem including SIFT [94], PolyPhen2 [95], MutationAssesor [96], FATHMM [97] and MutationTaster [98]. Although they were not intended for predicting driver mutations, these algorithms can be used to filter out variants that are unlikely to affect the structure and function of the protein, i.e. have more chance to be passenger mutations. For in-depth reviews of remaining challenges in the field of driver identification including prioritization of variants within the non-coding regions please refer to [99-100].

Some tools including FATHMM and MutationAssessor claim to assign higher functional impact scores to mutations occurring in driver genes. The former algorithm also has a special version in which a cancer-specific weighting scheme was incorporated to potentiate the functional analysis of driver mutations [101]. Similar approach is adopted in TransFIC software [84] where the scores obtained *via* PolyPhen2, SIFT and MutationAssessor are transformed in order to discriminate likely drivers from likely passengers. For each somatic mutation, its prediction score is compared with the distribution of scores for germline mutations located within functionally-related genes. Observed significant differences suggest that the mutation under study may be involved in cancer development.

Several algorithms like CHASM [102] and CanPredict [103] treat the differentiation between driver and passenger somatic missense mutations as a classification problem. The random forest classifier is trained to distinguish between driver mutations curated from COSMIC and passenger mutations generated according to background substitution frequencies. Each mutation is described by various features such as amino acid substitution properties, alignment-based estimates of conservation, predicted local structure, and etc.

In Funseq paper [104], instead of binary classification, variants are prioritized according to several criteria including occurrence of mutation in 1000 Genomes Project, breakage of transcription-factor binding site, location within gene under strong selection or in a hub gene and etc. Variants on the top of the list are more likely to be cancer drivers. An important feature of Funseq is its ability to prioritize mutations located in non-coding regions.

Still there is significant room for improvement of tools for driver mutation prediction. Assessment of several algorithms showed that no single method or combination of methods exceeded 81% accuracy [105].

CLINICAL INTERPRETATION AND PHARMACOGENOMICS

A reliable assessment of driver mutations, though it may help with identifying specific mechanisms of tumorigenesis, still may not have an immediate prognostic or treatment value. Currently, nearly all molecular therapies can directly target only driver genes with activating mutations (typically oncogenes, such as kinases). On the contrary, restoring loss-of-function alterations in many tumour suppressors requires other more complicated strategies like gene therapy [106], inhibition of a functionally connected gene [107], synthetic [108] and collateral [109] lethality. Review of promising approaches aimed at targeting tumor supressors therapeutically can be found in [110].

Results of high-throughput personalized genomic analyses show that conceivably actionable mutations are quite frequent. Jones with co-authors succeeded in identifying and linking somatic alterations in genes with potentially actionable consequences for 77% of cases [111]. This included associations with known therapies and current clinical trials. This estimate is very close to other results, [112] where the percentage of patients predicted to benefit from targeted agents (again, including clinical trials) was 73% of cases. However when considering only FDA-approved therapies (including drug repurposing), this value reduces to 40%, and considering only standard clinical guidelines, this percentage becomes 6%.

In 2014, the FDA included 10 new drugs and biologics in a list containing more than 165 pharmacogenetic labels for approved agents [3,113]. Only around 50 of them are related to oncology and clearly associated with efficacy, but some were for toxicity pharmacogenetic labels [114]. While the number of efficacy pharmacogenetic labels for the new drugs has doubled in the last four years (e.g. crizotinib, ceritinib for ALK rearrangements; bosutinib, omacetaxine and ponatinib for BCR-ABL fusion protein; dabrafenib, trametinib and vemurafenib for mutated BRAF), the number of toxicity markers (e.g. TPMT, DPD, G6PD and UGT1A1 deficiencies) has not grown so fast and generally corresponds to the older therapies [19].

The numbers presented above show a great promise for drug repurposing and illustrate the potential of precision oncology for a large number of patients whose tumors harbor potentially druggable alterations [19,115]. To promote the corresponding studies, several resources with information about clinically actionable somatic mutations were created.

One of the examples is a curated database TARGET (Tumor Alterations Relevant for Genomics-drivEn

Therapy) from Broad Institute, storing information for about 135 genes that may have therapeutic, prognostic and diagnostic implications [116]. It includes rationales behind each gene, the types of recurrent alterations that have clinical relevance in these genes, and the potential therapies. The Personalized Cancer Therapy (PCT) resource fromMD Anderson Cancer Center collects associations of genomic alterations with tumor development and growth, changes in response to therapy, availability of FDA-approved drugs, and investigational agents in clinical trials [117]. MyCancerGenome resource [www.mycancergenome.org] matches tumor mutations to targeted therapies including available clinical trials. The paper by Meric-Bernstam et al [117] provides a list of 120 potentially actionable genes for genomically informed therapy, the overlap with TARGET being approximately two thirds.

In addition to tumor-specific alterations which affect therapeutic efficacy, there are a number of germline genetic variants which can result in large interindividual differences in the pharmacokinetic profile of a drug [19]. Genetic alterations in genes responsible for drug metabolism and transport may lead to severe toxicities and should be taken into account by physicians for proper dose adjustment. Common examples of genotype effect and dose on toxicity include polymorphisms in TPMT and thiopurine drugs, UGT1A1 and irinotecan, DPD and 5-fluorouracil. [references]

So far, association between SNVs and drug response is the most studied variant in pharmacogenomics. However, other types of variants including CNAs, InDels and fusions can also guide therapy. A classic example is the use of trastuzumab for HER2-amplified/overexpressing breast cancer [118]. Examples of clinically relevant somatic fusions include EML4-ALK in non-small-cell lung cancer (sensitive to crizotinib [71]) and BCR-ABL fusions in chronic myelogenous leukemia (sensitive to imatinib [70]).

Researchers need to gain a deeper insight into the complex dynamics of subclonal architecture and its impact on disease outcome and prognosis [119]. Genomic stratification of cancers has usually relied on tumor profiling, so it reflects mutations present only in the majority of cancer cells. Intra-tumour clonal heterogeneity can restrict response to therapy, including the emergence of drug-resistant malignant cells and metastasis. Even minor subclones may be clinically relevant. Thus, it has been shown that patients with colorectal cancer harboring KRAS mutations in minor subclones were resistant to anti-EGFR antibodies.[reference]

In addition, resistance may develop in patients who initially responded to therapy. And possibility of evolution of different clones under the selective pressure of therapy has to be taken into account [120].

There are several resources containing information on gene-variant-drug relationships. One of the most authoritative sources is PharmGKB [121], which contains manually curated variant annotations, potentially clinically actionable gene-drug associations, genotype-based dosing guidelines, drug-centered pathways and other pharmacogenetic summaries for most FDA-approved drugs. Genes, for which there are known pharmacogenetic relationships are called VIPs, Very Important Pharmacogenes. Another valuable source of drug-gene interactions that also includes information about anti-neoplastic drugs is Drug-Gene Interaction database (DGIdb) [122]. It integrates data from several sources including PharmGKB, DrugBank, Therapeutic Target Database (TTD) and ClinicalTrials.gov and includes records about known drug targets as well as potentially druggable genes. An example of an NGS-oriented resource is PGMD, PharmacoGenomic Mutation Database from BioBase [123], a manually curated comprehensive collection of all genomic variants that have been reported to have a pharmacogenomic effect in human studies. Online access to PGMD is free for users from academic institutions.

TISSUE-SPECIFIC ANNOTATION, CANCER CELL LINES, EPIGENETICS

Tumor localization is known to be a strong factor for the observed molecular profile, restricting the application of drug therapy. For example, while BRAF V600-mutated melanomas are sensitive to vemurafenib, BRAF V600-mutated colorectal cancers may not be as sensitive [124, 125]. This problem leads to the necessity to take into account the tissue-specific information. One of the ways to use tissue 'prior' is to utilize data on high-throughput characterization of cancer cells, connecting genomic and transcriptomic alterations to drug response pharmacologic profiles. For some cancer localizations different high-throughput-based classification schemes (sometimes non-NGS) have been proposed leading to survival or treatment outcome predictions [126-129]. Ideally it would be perfect to obtain comprehensive omics data across large cohorts of patients but this approach is prohibitively expensive and limited in the scope of drugs that can be tested [130]. Instead, it is much more feasible to perform drug screening coupled with omics-analyses in cell cultures, given that cell line molecular profiles resemble corresponding primary tumours [131].

There are several papers devoted to combining patient data with molecular profiles and drug sensitivity of the cell lines, thereby predicting a possible response to the therapy. Geeleher with colleagues developed ridge regression models for the prediction of chemotherapeutic response in patients based on tumor gene expression and drug IC50 values from a large panel of cell lines [132]. In another study, the authors utilized partial least squares regression-based modeling framework in order to build drug sensitivity models for erlotinib or sorafenib [133].

The cell line panel was used as the training dataset, while the algorithm performance was evaluated using gene expression data from patients treated with the same drug. For a comparative analysis of 44 drug sensitivity prediction algorithms, please refer to [130].

There are several large-scale projects devoted to drug sensitivity of cancer cell lines. Genomics of Drug Sensitivity in Cancer Project [134] contains information about drug sensitivity to 138 anti-cancer therapeutics for more than 1000 human cancer cell lines. Cell lines are also characterized *via* transcriptome, genome-wide analysis of copy number gain/loss and sequences of 67 cancer-associated genes. Cancer Cell Line Encyclopedia [135] provides access to genomic, gene expression, chromosomal copy number and pharmacologic profiles of more than 1000 cell lines comprising 36 types of cancer. Cell-Miner resource [136] allows easy access to NCI-60 database compiled by the U.S. National Cancer Institute. This panel of 60 commonly used human cancer cell lines has been comprehensively characterized across various genomic, transcriptomic and pharmacologic platforms, including whole exome sequencing, several microarray platforms, and sensitivity to 20 000 compounds including 102 FDA-approved drugs.

We speculate that the broad employment of *in silico* analysis of cell lines data may be a promising option for personalizing drug treatment. It is cheaper and faster compared to performing *in vitro* experiments like mouse xenograft models and allows screening hundreds of drugs in parallel, at the same time taking into account specific molecular profiles. However, cancer cell lines have several drawbacks including the difficulties modeling tumor heterogeneity and microenvironment [137]. Also, recent studies raise important questions regarding the poor reproducibility of results [138] and uncertain consistency across different sources of pharmacological data [139, 140].

Another aspect that should be taken into account is the impact of epigenetic mechanisms upon tumorigenesis such as inactivation of tumor suppressors *via* promoter methylation or histone modifications [141]. Although epigenetic data is rarely available for cancer cell lines, it may still be important for the selection of therapy, such as in case of colorectal cancer [142]. An interesting resource is dbEM database which compiles information about gene essentiality, mutation, copy number variation and expression level of epigenetic proteins from thousands of tumors and cancer cell lines [143]. For in-depth discussion of the possible role of epigenetic abnormalities in cancers, please refer to [144, 145].

SYSTEMS BIOLOGY AND DATA INTEGRATION

For the selection of optimal pharmacotherapy for an individual patient, it is necessary to understand precisely which molecular mechanisms drive the tumor progression. This problem can be addressed with a systems biology approach - how to interpret expression and mutation data and take them to a higher level of understanding. Here we will briefly describe basic applications of systems biology and data integration, while more details on this topic could be found in other publications [146-148].

BIOLOGICAL PATHWAY RESOURCES

The basis for the systems biology analysis is the biological knowledge represented in the form of relations between various molecular entities: genes/proteins, complexes, small molecules and etc. In its simplest form, this information can be represented as a collection of genesets, i.e. groups of functionally related genes. The most commonly used genesets can be obtained either *via* Gene Ontology annotations [149] or *via* MSigDB signatures [150]. More sophisticated expert knowledge is represented in the form of signalling and metabolic pathways describing specific biochemical processes. The most commonly used public pathway resources include KEGG PATHWAY [151], BioCarta [152] and Reactome [153]. Several databases accumulate information from a number of other pathway resources: ConsensusPathDB [154], PathwayCommons [155]. Finally, biological knowledge can be represented not as a set of separate pathways (whose boundaries are set more or less arbitrarily), but rather as a global network containing tens of thousands of entities interconnected by various types of physical and genetic interactions. Examples of such resources are STRING [156] and BioGRID [157].

The drug action/metabolism pathways describing pharmacokinetics and pharmacodynamics of a drug with potential pharmacogenetic associations are of particular value for the study of personalized medicine. This category of pathways can be downloaded from PharmGKB [121] or The Small Molecule Pathway Database [158]. Integration of mutation calls with drug pathways may identify proteins that can be targeted by the earlier approved drugs.

It should be noted that relatively few public resources provide cancer-specific pathways. Pathway Interaction Database [159] contains publicly available collection of curated and peer-reviewed pathways implicated in cancer. However, this database has not been updated since 2012. The Molecular Signatures Database stores hundreds of gene signatures which are often dysregulated in cancer. However, a majority of these signatures were generated directly from transcriptomics experiments rather than created by experts and hence may be unreliable. Network of Cancer Genes [160], although it is not a pathway resource in the literal sense, reports information on interactions, functions and expression of approximately 2000 of known and candidate cancer genes and oncomiRs.

A particularly valuable cancer-specific database

is ACSN, Atlas of Cancer Signalling Networks [161]. This resource aims to provide comprehensive maps of signalling and regulatory molecular processes that are frequently deregulated during cancerogenesis underlying the cancer hallmarks [162]. The key idea of ACSN is to consider cancerogenesis at several hierarchical levels: from bird-eye view maps of the 5 biological processes such as cell cycle or DNA repair through 52 detailed functional modules down to the seamless global network with thousands of molecular interactions. The whole cancer signalling network can be browsed with Google Maps interface allowing various zoom levels.

Several pathway resources provide a web interface allowing to overlay researcher's own data on available pathways. For example KEGGViewer [163] and Reactome [153] provide tools for coloring pathways according to expression data. ACSN resource allows users to overlay any expression, copy-number and mutation data on cancer maps, facilitating its biological interpretation.

Standalone pathway applications usually provide richer functionality and enable more sophisticated types of system-biological analyses using pathways and networks. Among the free standalone applications the most frequently used tool is Cytoscape [164]. Cytoscape offers rich opportunities for visualization of biological pathways and networks and integrates them with any data attributes, including gene expression and copy-number variations. Cytoscape supports many formats for data exchange and can be extended by more than 200 community-developed plugins, covering a variety of systems biology algorithms (a good intro can be found in [165]). Other applications for pathway visualization include PathVisio (integrated with WikiPathways database) [166] and GenMAPP [167].

Commercial pathway packages, such as Ingenuity® Pathway Analysis [27], MetaCore™ [168] and Pathway Studio® [169] come with comprehensive and carefully curated proprietary molecular databases and produce visually appealing networks. All these products include support for functional analysis of NGS data.

FUNCTIONAL INTERPRETATION OF DATA

One of the priorities in selecting personalized anticancer pharmacotherapy is to understand which specific signalling pathways are perturbed in an individual patient. Over the last decade, considerable efforts were undertaken in this direction, stimulated by the successes of microarray technology. The most simple approach (although the most popular one), the so-called over-representation analysis, consists of identifying genes with significantly altered expression in tumors followed by finding predefined genesets/pathways where the observed fraction of altered genes differs from the expected value. Discovery of such overrepresented genesets allows researcher to interpret expression data for the individual

patient in terms of pathways.

Many tools have been developed for performing overrepresentation analysis including DAVID [170] and WebGestalt [171]. Most methods are commonly based on hypergeometric distribution (Fisher's exact test) and differ very slightly from each other [146]. The drawback of this approach is the necessity to determine differentially expressed genes, which is not a trivial task for NGS data [172]. It is possible to apply the overrepresentation analysis for other types of variations in order to identify pathways enriched by mutated genes. However, in this case, the results will be noisy due to passenger mutations, which comprise the majority of somatic genetic variations.

A more advanced class of methods, the so-called functional class scoring, works directly with all genes measured in an experiment that are ranked, for example, by the strength of differential expression. If the rank distribution of genes within a specific geneset significantly differs from the background, then this geneset is somehow activated. This approach provides greater sensitivity in detecting small, though coordinated, expression changes of functionally related genes. The most commonly used algorithm implementing functional class scoring is the gene set enrichment analysis, GSEA [173]. There are many extensions and improvements of this classical algorithm including single-sample oriented analysis like GSVA [174] and ssGSEA [175].

A natural extension of the above methods is to utilize additional biological knowledge presented in the form of relations between entities in the pathway. Several algorithms have been developed implementing this idea. One popular method, SPIA, combines standard overrepresentation test with a measure of the actual perturbation on a given pathway taking into account relative gene locations [176]. Another method, DEAP, identifies those pathways where observed expression data is better "explained" by activatory/inhibitory relations between genes [177]. Pathway topology-based algorithms claim to have better specificity and more sensitivity compared to classical approaches.

However, when used for NGS data, most of the traditional approaches for functional interpretation are prone to potential biases and should be applied with care. Given that the genetic alterations occur evenly across the genome, long genes tend to harbor more mutations. Hence, the results of over-representation analysis for the list of mutated genes will be biased for pathways containing longer genes than other pathways [178]. A similar effect is present in the analysis of differential expression of RNASeq data: longer transcripts generate a greater number of reads and are more likely to be detected as differentially expressed compared with their short counterparts [179]. Several enrichment-based algorithms explicitly take into account this long-gene effect, including GOSeq [180], SeqGSEA [181], GSVA [174], and GOglm [182].

DATA INTEGRATION

The data integration, as a union of the results obtained by various omics technologies, is of special importance when dealing with cancer data. Biology of cancer cells is extremely complex with alterations occurring on (epi)genomic, transcriptomic, proteomic and metabolomic levels. Hence, in order to improve statistical and interpretative power and obtain reliable view of an individual's tumor biology, it is necessary to sum up the maximum possible number of sources of information, each capturing a different aspect of cancerogenesis.

There is no single standard approach to data integration. While several algorithms have been developed such as GSAA [183], iCluster+ [184] and GSOA [185], most of them are designed to deal with a cohort of samples and require large training sets. One of the best known methods of data integration is PARADIGM [186, 187] which utilizes CNA and gene expression in order to infer patient-specific genetic activities. An example of a true single-sample approach using multiple sources of data is PHIAL [116] - an algorithm for annotation and ranking a patient's somatic alterations on the basis of their clinical and biological relevance. This approach takes into account SNV data, CNAs, and chromosomal rearrangement, as well as intrasample pathway structure: a mutation located within a gene connected to other gene with a known actionable alteration receives higher score.

Data integration can also be performed by visualizing results of various analyses on the same plot. There are several ways to depict multidimensional oncogenomics data such as matrix heatmaps, genomic coordinates, and networks (see [188] for a comprehensive review). These plots can be built using various standalone applications and websites such as GItools [189], IntOGen [190], cBioportal[92], Cytoscape [164]. A commonly used option is circos plots [191], where the genomic coordinates of all chromosomes are represented in a circular layout and where additional data tracks may include mutation pattern, CNAs, genomic rearrangements etc.

Methods for identification of activated pathways can also be considered as an approach to data integration since they allow aggregating various types of molecular events across several genes in the common feature space, simplifying data interpretation and gaining insight into the biological system [147]. For example, consider a specific pathway which has been predicted as activated according to transcriptomics analysis and also contains a mutated transcription factor (for example, MYC) "explaining" observed changes in expression. This finding directly points to potential causal mechanisms for tumor progression and can give clues to the needed therapy.

CONCLUSIONS

Since introduction of the first predictive cancer biomarkers to the clinics, much progress has been made in the area of Precision Oncology. The number of genes associated with therapy choice has grown significantly, and corresponding variation detection has started to require NGS application. The «Exceptional responders» approach has been developed to find new predictive biomarkers. The patients with the best response to the tumor-specific therapy are studied thoroughly using the methods of NGS in order to find characteristic molecular features.

However, using standard clinical trials design, it may be difficult to confirm the clinical importance of these new genomic variations because of their low frequency. To overcome this issue, new "basket" type of clinical trials with patients stratified on the base of tumor molecular profile only is being developed.

As the cost of sequencing decreases, another approach becomes more popular. In this approach, the molecular profile of each patient is studied as completely as possible, taking into account the specific properties of the studied tissues. The systems biology analysis and integration of different kinds of NGS data play a critical role for detection of the most probable targets for the personalized therapy. In this article, we reviewed some examples of the corresponding case-studies, the general approach to this kind of the data interpretation, and the specific instruments that can be used. Despite the fact that the existing examples are quite promising, further development and verification of standards in NGS data processing for Precision Oncology is still necessary.

Abbreviations

aCGH, Comparative Genomic Hybridization; ALL, acute lymphoblastic leukemia; AML, Acute myeloid leukemia; CISH, chromogenic In situ hybridization; CMI, Caris Molecular Intelligence platform; CNAs, copy number alterations; DPD, dihydropyrimidine dehydrogenase; CpG (sites), regions of DNA where a cytosine nucleotide occurs next to a guanine nucleotide separated by only one phosphate; FDA, Food and Drug Administration; FFPE, fresh-frozen paraffin embedded block; FISH, Fluorescence in situ hybridization; G6PD, Glucose-6-phosphate dehydrogenase; IHC, immunohistochemistry; InDels, insertions and deletions; LDH, lactate dehydrogenase; LOH, Loss of heterozygosity; MDS, myelodysplastic syndromes; MPN, Myeloproliferative neoplasms; mtDNA, mitochondrial DNA; NCI, National Cancer Institute; NGS, new generation sequencing; NIH, National Institutes of Health; NSCLC, Non-small-cell lung carcinoma; PSA, prostate-specific antigen; RNA-Seq, RNA sequencing; RT-PCR, reverse transcription polymerase chain reaction; SNVs. single nucleotide variations; TPMT, thiopurine methyltransferase; WES, whole exome sequencing; WGS, whole genome sequencing.

ACKNOWLEDGMENTS

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

GRANT SUPPORT

REFERENCES

1. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 2000; 100:57–70.

2. De Palma M, Hanahan D. The biology of personalized cancer medicine: facing individual complexities underlying hallmark capabilities. Mol Oncol 2012; 6:111–27.

3. Center For Drug Evaluation. Genomics - Table of Pharmacogenomic Biomarkers in Drug Labeling. [cited 2015 Oct 2]; Available from: http://www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics/ucm083378.html

4. Yeh P, Chen H, Andrews J, Naser R, Pao W, Horn L. DNA-Mutation Inventory to Refine and Enhance Cancer Treatment (DIRECT): a catalog of clinically relevant cancer mutations to enable genome-directed anticancer therapy. Clin Cancer Res 2013; 19:1894–901.

5. McCann GA, Eisenhauer EL. Hereditary cancer syndromes with high risk of endometrial and ovarian cancer: surgical options for personalized care. J Surg Oncol 2015; 111:118–24.

6. Powell CB. Clinical management of patients at inherited risk for gynecologic cancer. Curr Opin Obstet Gynecol 2015; 27:14–22.

7. Weiss G, Glen W, Brandi H, Robert W, Ashish S, Susan G, Robert P, David M, Scott M, Eric T, David L, Vivek K. Evaluation and comparison of two commercially available targeted next-generation sequencing platforms to assist oncology decision making. Onco Targets Ther 2015; :959.

8. Shen T, Pajaro-Van de Stadt SH, Yeat NC, Lin JC-H. Clinical applications of next generation sequencing in cancer: from panels, to exomes, to genomes. Front Genet 2015; 6:215.

9. Beltran H, Yelensky R, Frampton GM, Park K, Downing SR, MacDonald TY, Jarosz M, Lipson D, Tagawa ST, Nanus DM, Stephens PJ, Mosquera JM, Cronin MT, et al. Targeted next-generation sequencing of advanced prostate cancer identifies potential therapeutic targets and disease heterogeneity. Eur Urol 2013; 63:920–6.

10. Zhao X, Wang A, Walter V, Patel NM, Eberhard DA, Hayward MC, Salazar AH, Jo H, Soloway MG, Wilkerson MD, Parker JS, Yin X, Zhang G, et al. Combined Targeted DNA Sequencing in Non-Small Cell Lung Cancer (NSCLC) Using UNCseq and NGScopy, and RNA Sequencing Using UNCqeR for the Detection of Genetic Aberrations in NSCLC. PLoS One 2015; 10:e0129280.

11. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O'Daniel JM, Ormond KE, Rehm HL, Watson MS, Williams MS, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med 2013; 15:565–74.

12. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. Science 2013; 339:1546–58.

13. Mutz K-O, Heilkenbrinker A, Lönne M, Walter J-G, Stahl F. Transcriptome analysis using next-generation sequencing. Curr Opin Biotechnol 2013; 24:22–30.

14. Stirzaker C, Zotenko E, Song JZ, Qu W, Nair SS, Locke WJ, Stone A, Armstong NJ, Robinson MD, Dobrovic A, Avery-Kiejda KA, Peters KM, French JD, et al. Methylome sequencing in triple-negative breast cancer reveals distinct methylation clusters with prognostic value. Nat Commun 2015; 6:5899.

15. Bennett NC, Farah CS. Next-generation sequencing in clinical oncology: next steps towards clinical validation. Cancers 2014; 6:2296–312.

16. Gagan J, Van Allen EM. Next-generation sequencing to guide cancer therapy. Genome Med 2015; 7:80.

17. Sheridan C. Milestone approval lifts Illumina's NGS from research into clinic. Nat Biotechnol 2014; 32:111–2.

18. Shrager J, Tenenbaum JM. Rapid learning for precision oncology. Nat Rev Clin Oncol 2014; 11:109–18.

19. Rodríguez-Antona C, Taron M. Pharmacogenomic biomarkers for personalized cancer treatment. J Intern Med 2015; 277:201–17.

20. Gray SW, Cronin A, Bair E, Lindeman N, Viswanath V, Janeway KA. Marketing of personalized cancer care on the web: an analysis of Internet websites. J Natl Cancer Inst [Internet] 2015; 107. Available from: http://dx.doi.org/10.1093/jnci/djv030

21. Pham-Ledard A, Cowppli-Bony A, Doussau A, Prochazkova-Carlotti M, Laharanne E, Jouary T, Belaud-Rotureau M-A, Vergier B, Merlio J-P, Beylot-Barry M. Diagnostic and prognostic value of BCL2 rearrangement in 53 patients with follicular lymphoma presenting as primary skin lesions. Am J Clin Pathol 2015; 143:362–73.

22. Ito T, Hamasaki M, Matsumoto S, Hiroshima K, Tsujimura T, Kawai T, Shimao Y, Marutsuka K, Moriguchi S, Maruyama R, Miyamoto S, Nabeshima K. p16/CDKN2A FISH in Differentiation of Diffuse Malignant Peritoneal

Mesothelioma From Mesothelial Hyperplasia and Epithelial Ovarian Cancer. Am J Clin Pathol 2015; 143:830–8.

23. Mertens F, Johansson B, Fioretos T, Mitelman F. The emerging complexity of gene fusions in cancer. Nat Rev Cancer 2015; 15:371–81.

24. Von Hoff DD, Stephenson JJ Jr, Rosen P, Loesch DM, Borad MJ, Anthony S, Jameson G, Brown S, Cantafio N, Richards DA, Fitch TR, Wasserman E, Fernandez C, et al. Pilot study using molecular profiling of patients' tumors to find potential targets and select treatments for their refractory cancers. J Clin Oncol 2010; 28:4877–83.

25. Van Allen EM, Wagle N, Levy MA. Clinical analysis and interpretation of cancer genome data. J Clin Oncol 2013; 31:1825–33.

26. Jones SJ, Laskin J, Li YY, Griffith OL, An J, Bilenky M, Butterfield YS, Cezard T, Chuah E, Corbett R, Fejes AP, Griffith M, Yee J, et al. Evolution of an adenocarcinoma in response to selection by targeted kinase inhibitors. Genome Biol 2010; 11:R82.

27. Ingenuity IPA - Integrate and understand complex 'omics data [Internet]. Ingenuity [cited 2015 Oct 2]; Available from: http://www.ingenuity.com/products/ipa

28. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000; 28:27–30.

29. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, Djoumbou Y, Eisner R, Guo AC, et al. DrugBank 3.0: a comprehensive resource for "omics" research on drugs. Nucleic Acids Res 2011; 39:D1035–41.

30. Welch JS, Westervelt P, Ding L, Larson DE, Klco JM, Kulkarni S, Wallis J, Chen K, Payton JE, Fulton RS, Veizer J, Schmidt H, Vickery TL, et al. Use of whole-genome sequencing to diagnose a cryptic fusion oncogene. JAMA 2011; 305:1577–84.

31. Kolata G. In Gene Sequencing Treatment for Leukemia, Glimpses of the Future. NY Times [Internet] 2012 [cited 2015 Oct 2]; Available from: http://www.nytimes.com/2012/07/08/health/in-gene-sequencing-treatment-for-leukemia-glimpses-of-the-future.html

32. Russell K, Shunyakov L, Dicke KA, Maney T, Voss A. A practical approach to aid physician interpretation of clinically actionable predictive biomarker results in a multi-platform tumor profiling service. Front Pharmacol 2014; 5:76.

33. Le Tourneau C, Paoletti X, Servant N, Bièche I, Gentien D, Rio Frio T, Vincent-Salomon A, Servois V, Romejon J, Mariani O, Bernard V, Huppe P, Pierron G, et al. Randomised proof-of-concept phase II trial comparing targeted therapy based on tumour molecular profiling vs conventional therapy in patients with refractory cancer: results of the feasibility part of the SHIVA trial. Br J Cancer 2014; 111:17–24.

34. Servant N, Roméjon J, Gestraud P, La Rosa P, Lucotte G, Lair S, Bernard V, Zeitouni B, Coffin F, Jules-Clément G,

Yvon F, Lermine A, Poullet P, et al. Bioinformatics for precision medicine in oncology: principles and application to the SHIVA clinical trial. Front Genet 2014; 5:152.

35. DCTD — Major Initiatives [Internet]. [cited 2015 Oct 2]; Available from: http://dctd.cancer.gov/MajorInitiatives/NCI-sponsored_trials_in_precision_medicine.htm

36. Beltran H, Eng K, Mosquera JM, Sigaras A, Romanel A, Rennert H, Kossai M, Pauli C, Faltas B, Fontugne J, Park K, Banfelder J, Prandi D, et al. Whole-Exome Sequencing of Metastatic Cancer and Biomarkers of Treatment Response. JAMA Oncol 2015; 1:466–74.

37. Zhao Y, Polley EC, Li M-C, Lih C-J, Palmisano A, Sims DJ, Rubinstein LV, Conley BA, Chen AP, Williams PM, Kummar S, Doroshow JH, Simon RM. GeneMed: An Informatics Hub for the Coordination of Next-Generation Sequencing Studies that Support Precision Oncology Clinical Trials. Cancer Inform 2015; 14:45–55.

38. Rodon J, Soria JC, Berger R, Batist G, Tsimberidou A, Bresson C, Lee JJ, Rubin E, Onn A, Schilsky RL, Miller WH, Eggermont AM, Mendelsohn J, et al. Challenges in initiating and conducting personalized cancer therapy trials: perspectives from WINTHER, a Worldwide Innovative Network (WIN) Consortium trial. Ann Oncol 2015; 26:1791–8.

39. DCTD — Major Initiatives [Internet]. [cited 2015 Oct 2]; Available from: http://dctd.cancer.gov/MajorInitiatives/NCI-sponsored_trials_in_precision_medicine.htm

40. Brower V. NCI-MATCH pairs tumor mutations with matching drugs. Nat Biotechnol 2015; 33:790–1.

41. Do K, O'Sullivan Coyne G, Chen AP. An overview of the NCI precision medicine trials-NCI MATCH and MPACT. Chin Clin Oncol 2015; 4:31.

42. Dabrowski A, Chwiećko M. Oxygen radicals mediate depletion of pancreatic sulfhydryl compounds in rats with cerulein-induced acute pancreatitis. Digestion 1990; 47:15–9.

43. Wagle N, Grabiner BC, Van Allen EM, Hodis E, Jacobus S, Supko JG, Stewart M, Choueiri TK, Gandhi L, Cleary JM, Elfiky AA, Taplin ME, Stack EC, et al. Activating mTOR mutations in a patient with an extraordinary response on a phase I trial of everolimus and pazopanib. Cancer Discov 2014; 4:546–53.

44. Kitano H. Systems biology: a brief overview. Science 2002; 295:1662–4.

45. Ideker T, Galitski T, Hood L. A new approach to decoding life: systems biology. Annu Rev Genomics Hum Genet 2001; 2:343–72.

46. Werner HMJ, Mills GB, Ram PT. Cancer Systems Biology: a peek into the future of patient care? Nat Rev Clin Oncol 2014; 11:167–76.

47. Wolkenhauer O, Auffray C, Jaster R, Steinhoff G, Dammann O. The road from systems biology to systems medicine. Pediatr Res 2013; 73:502–7.

48. Lili LN, Matyunina LV, Walker LD, Daneker GW,

McDonald JF. Evidence for the importance of personalized molecular profiling in pancreatic cancer. Pancreas 2014; 43:198–211.

49. Barillot E, Calzone L, Hupe P, Vert J-P, Zinovyev A. Computational systems biology of cancer. CRC Press; 2012.

50. Garcia-Murillas I, Schiavon G, Weigelt B, Ng C, Hrebien S, Cutts RJ, Cheang M, Osin P, Nerurkar A, Kozarewa I, Garrido JA, Dowsett M, Reis-Filho JS, et al. Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. Sci Transl Med 2015; 7:302ra133.

51. Ramsköld D, Luo S, Wang Y-C, Li R, Deng Q, Faridani OR, Daniels GA, Khrebtukova I, Loring JF, Laurent LC, Schroth GP, Sandberg R. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol 2012; 30:777–82.

52. Graw S, Meier R, Minn K, Bloomer C, Godwin AK, Fridley B, Vlad A, Beyerlein P, Chien J. Robust gene expression and mutation analyses of RNA-sequencing of formalin-fixed diagnostic tumor samples. Sci Rep 2015; 5:12335.

53. Tuononen K, Mäki-Nevala S, Sarhadi VK, Wirtanen A, Rönty M, Salmenkivi K, Andrews JM, Telaranta-Keerie AI, Hannula S, Lagström S, Ellonen P, Knuuttila A, Knuutila S. Comparison of targeted next-generation sequencing (NGS) and real-time PCR in the detection of EGFR, KRAS, and BRAF mutations on formalin-fixed, paraffin-embedded tumor material of non-small cell lung carcinoma-superiority of NGS. Genes Chromosomes Cancer 2013; 52:503–11.

54. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature 2014; 511:543–50.

55. Krauthammer M, Kong Y, Ha BH, Evans P, Bacchiocchi A, McCusker JP, Cheng E, Davis MJ, Goh G, Choi M, Ariyan S, Narayan D, Dutton-Regester K, et al. Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. Nat Genet 2012; 44:1006–14.

56. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet 2010; 11:31–46.

57. Xuan J, Yu Y, Qing T, Guo L, Shi L. Next-generation sequencing in the clinic: promises and challenges. Cancer Lett 2013; 340:284–95.

58. Sboner A, Mu XJ, Greenbaum D, Auerbach RK, Gerstein MB. The real cost of sequencing: higher than you think! Genome Biol 2011; 12:125.

59. Ding L, Wendl MC, McMichael JF, Raphael BJ. Expanding the computational toolbox for mining cancer genomes. Nat Rev Genet 2014; 15:556–70.

60. Ewing AD, Houlahan KE, Hu Y, Ellrott K, Caloian C, Yamaguchi TN, Bare JC, P'ng C, Waggott D, Sabelnykova VY, ICGC-TCGA DREAM Somatic Mutation Calling Challenge participants, Kellen MR, Norman TC, et al. Combining tumor genome simulation with crowdsourcing to benchmark somatic single-nucleotide-variant detection. Nat Methods 2015; 12:623–30.

61. Roberts ND, Kortschak RD, Parker WT, Schreiber AW, Branford S, Scott HS, Glonek G, Adelson DL. A comparative analysis of algorithms for somatic SNV detection in cancer. Bioinformatics 2013; 29:2223–30.

62. Wilkerson MD, Cabanski CR, Sun W, Hoadley KA, Walter V, Mose LE, Troester MA, Hammerman PS, Parker JS, Perou CM, Hayes DN. Integrated RNA and DNA sequencing improves mutation detection in low purity tumors. Nucleic Acids Res 2014; 42:e107.

63. Fang H, Wu Y, Narzisi G, O'Rawe JA, Barrón LTJ, Rosenbaum J, Ronemus M, Iossifov I, Schatz MC, Lyon GJ. Reducing INDEL calling errors in whole genome and exome sequencing data. Genome Med 2014; 6:89.

64. Li H. Toward better understanding of artifacts in variant calling from high-coverage samples. Bioinformatics 2014; 30:2843–51.

65. Mose LE, Wilkerson MD, Hayes DN, Perou CM, Parker JS. ABRA: improved coding indel detection via assembly-based realignment. Bioinformatics 2014; 30:2813–5.

66. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010; 20:1297–303.

67. Alkodsi A, Louhimo R, Hautaniemi S. Comparative analysis of methods for identifying somatic copy number alterations from deep sequencing data. Brief Bioinform 2015; 16:242–54.

68. Liu B, Morrison CD, Johnson CS, Trump DL, Qin M, Conroy JC, Wang J, Liu S. Computational methods for detecting copy number variations in cancer genome using next generation sequencing: principles and challenges. Oncotarget 2013; 4:1868–81. doi: 10.18632/ oncotarget.1537.

69. Graham RP, Jin L, Knutson DL, Kloft-Nelson SM, Greipp PT, Waldburger N, Roessler S, Longerich T, Roberts LR, Oliveira AM, Halling KC, Schirmacher P, Torbenson MS. DNAJB1-PRKACA is specific for fibrolamellar carcinoma. Mod Pathol 2015; 28:822–9.

70. An X, Tiwari AK, Sun Y, Ding P-R, Ashby CR Jr, Chen Z-S. BCR-ABL tyrosine kinase inhibitors in the treatment of Philadelphia chromosome positive chronic myeloid leukemia: a review. Leuk Res 2010; 34:1255–68.

71. Kwak EL, Bang Y-J, Camidge DR, Shaw AT, Solomon B, Maki RG, Ou S-HI, Dezube BJ, Jänne PA, Costa DB, Varella-Garcia M, Kim W-H, Lynch TJ, et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. N Engl J Med 2010; 363:1693–703.

72. Yoshihara K, Wang Q, Torres-Garcia W, Zheng S, Vegesna R, Kim H, Verhaak RGW. The landscape and therapeutic relevance of cancer-associated transcript fusions. Oncogene 2015; 34:4845–54.

73. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L.

Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 2012; 7:562–78.

74. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014; 15:550.

75. Hardcastle TJ, Kelly KA. baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics 2010; 11:422.

76. Laird PW, Jaenisch R. DNA methylation and cancer. Hum Mol Genet 1994; 3 Spec No:1487–95.

77. Stirzaker C, Taberlay PC, Statham AL, Clark SJ. Mining cancer methylomes: prospects and challenges. Trends Genet 2014; 30:75–84.

78. Rohlin A, Wernersson J, Engwall Y, Wiklund L, Björk J, Nordling M. Parallel sequencing used in detection of mosaic mutations: comparison with four diagnostic DNA screening techniques. Hum Mutat 2009; 30:1012–20.

79. Biesecker LG, Spinner NB. A genomic view of mosaicism and human disease. Nat Rev Genet 2013; 14:307–20.

80. Machiela MJ, Chanock SJ. Detectable clonal mosaicism in the human genome. Semin Hematol 2013; 50:348–59.

81. Friedman E, Efrat N, Soussan-Gutman L, Dvir A, Kaplan Y, Ekstein T, Nykamp K, Powers M, Rabideau M, Sorenson J, Topper S. Low-level constitutional mosaicism of a de novoBRCA1 gene mutation. Br J Cancer 2015; 112:765–8.

82. Yamaguchi K, Komura M, Yamaguchi R, Imoto S, Shimizu E, Kasuya S, Shibuya T, Hatakeyama S, Takahashi N, Ikenoue T, Hata K, Tsurita G, Shinozaki M, et al. Detection of APC mosaicism by next-generation sequencing in an FAP patient. J Hum Genet 2015; 60:227–31.

83. Pasmant E, Parfait B, Luscan A, Goussard P, Briand-Suleau A, Laurendeau I, Fouveaut C, Leroy C, Montadert A, Wolkenstein P, Vidaud M, Vidaud D. Neurofibromatosis type 1 molecular diagnosis: what can NGS do for you when you have a large gene with loss of function mutations? Eur J Hum Genet 2015; 23:596–601.

84. Krishnan VG, Ng PC. Predicting cancer drivers: are we there yet? Genome Med 2012; 4:88.

85. Dancey JE, Bedard PL, Onetto N, Hudson TJ. The genetic basis for cancer treatment decisions. Cell 2012; 148:409–20.

86. Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, Chen S, Karchin R, Kinzler KW, Vogelstein B, Nowak MA. Accumulation of driver and passenger mutations during tumor progression. Proc Natl Acad Sci U S A 2010; 107:18545–50.

87. McFarland CD, Korolev KS, Kryukov GV, Sunyaev SR, Mirny LA. Impact of deleterious passenger mutations on cancer progression. Proc Natl Acad Sci U S A 2013; 110:2910–5.

88. Tamborero D, Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Kandoth C, Reimand J, Lawrence MS, Getz G, Bader GD, Ding L, Lopez-Bigas N. Comprehensive identification of mutational cancer driver genes across 12 tumor types. Sci Rep 2013; 3:2650.

89. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, Meyerson M, Gabriel SB, Lander ES, Getz G. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 2014; 505:495–501.

90. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, Ding M, Bamford S, Cole C, Ward S, Kok CY, Jia M, De T, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res 2015; 43:D805–11.

91. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 2013; 45:1113–20.

92. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012; 2:401–4.

93. Cline MS, Craft B, Swatloski T, Goldman M, Ma S, Haussler D, Zhu J. Exploring TCGA Pan-Cancer data at the UCSC Cancer Genomics Browser. Sci Rep 2013; 3:2652.

94. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res 2001; 11:863–74.

95. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods 2010; 7:248–9.

96. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res 2011; 39:e118.

97. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GLA, Edwards KJ, Day INM, Gaunt TR. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum Mutat 2013; 34:57–65.

98. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 2010; 7:575–6.

99. Pon JR, Marra MA. Driver and passenger mutations in cancer. Annu Rev Pathol 2015; 10:25–50.

100. Marx V. Cancer genomes: discerning drivers from passengers. Nat Methods 2014; 11:375–9.

101. Shihab HA, Gough J, Cooper DN, Day INM, Gaunt TR. Predicting the functional consequences of cancer-associated amino acid substitutions. Bioinformatics 2013; 29:1504–10.

102. Douville C, Carter H, Kim R, Niknafs N, Diekhans M, Stenson PD, Cooper DN, Ryan M, Karchin R. CRAVAT: cancer-related analysis of variants toolkit. Bioinformatics 2013; 29:647–8.

103. Kaminker JS, Zhang Y, Watanabe C, Zhang Z. CanPredict: a computational tool for predicting cancer-associated

missense mutations. Nucleic Acids Res 2007; 35:W595–8.

104. Khurana E, Fu Y, Colonna V, Mu XJ, Kang HM, Lappalainen T, Sboner A, Lochovsky L, Chen J, Harmanci A, Das J, Abyzov A, Balasubramanian S, et al. Integrative annotation of variants from 1092 humans: application to cancer genomics. Science 2013; 342:1235587.

105. Gnad F, Baucom A, Mukhyala K, Manning G, Zhang Z. Assessment of computational methods for predicting the effects of missense mutations in human cancers. BMC Genomics 2013; 14 Suppl 3:S7.

106. Hong B, van den Heuvel APJ, Prabhu VV, Zhang S, El-Deiry WS. Targeting tumor suppressor p53 for cancer therapy: strategies, challenges and opportunities. Curr Drug Targets 2014; 15:80–9.

107. Nag S, Zhang X, Srivenugopal KS, Wang M-H, Wang W, Zhang R. Targeting MDM2-p53 interaction for cancer therapy: are we there yet? Curr Med Chem 2014; 21:553–74.

108. Jackson RA, Chen ES. Synthetic lethal approaches for assessing combinatorial efficacy of chemotherapeutic drugs. Pharmacol Ther [Internet] 2016; Available from: http://dx.doi.org/10.1016/j.pharmthera.2016.01.014

109. Muller FL, Aquilanti EA, DePinho RA. Collateral Lethality: A new therapeutic strategy in oncology. Trends Cancer Res 2015; 1:161–73.

110. Morris LGT, Chan TA. Therapeutic targeting of tumor suppressor genes. Cancer 2015; 121:1357–68.

111. Jones S, Anagnostou V, Lytle K, Parpart-Li S, Nesselbush M, Riley DR, Shukla M, Chesnick B, Kadan M, Papp E, Galens KG, Murphy D, Zhang T, et al. Personalized genomic analyses for cancer mutation discovery and interpretation. Sci Transl Med 2015; 7:283ra53.

112. Rubio-Perez C, Tamborero D, Schroeder MP, Antolín AA, Deu-Pons J, Perez-Llamas C, Mestres J, Gonzalez-Perez A, Lopez-Bigas N. In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. Cancer Cell 2015; 27:382–96.

113. American Society of Clinical Oncology. The state of cancer care in america, 2015: a report by the american society of clinical oncology. J Oncol Pract 2015; 11:79–113.

114. Wang L, Liu H, Chute CG, Zhu Q. Cancer based pharmacogenomics network supported with scientific evidences: from the view of drug repurposing. BioData Min 2015; 8:9.

115. Martinez-Ledesma E, de Groot JF, Verhaak RGW. Seek and destroy: relating cancer drivers to therapies. Cancer Cell 2015; 27:319–21.

116. Van Allen EM, Wagle N, Stojanov P, Perrin DL, Cibulskis K, Marlow S, Jane-Valbuena J, Friedrich DC, Kryukov G, Carter SL, McKenna A, Sivachenko A, Rosenberg M, et al. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. Nat Med 2014; 20:682–8.

117. Meric-Bernstam F, Johnson A, Holla V, Bailey AM, Brusco L, Chen K, Routbort M, Patel KP, Zeng J, Kopetz S, Davies MA, Piha-Paul SA, Hong DS, et al. A decision support framework for genomically informed investigational cancer therapy. J Natl Cancer Inst [Internet] 2015; 107. Available from: http://dx.doi.org/10.1093/jnci/djv098

118. Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE Jr, Davidson NE, Tan-Chiu E, Martino S, Paik S, Kaufman PA, Swain SM, Pisansky TM, Fehrenbacher L, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. N Engl J Med 2005; 353:1673–84.

119. Sutton L-A, Rosenquist R. Clonal evolution in chronic lymphocytic leukemia: impact of subclonality on disease progression. Expert Rev Hematol 2015; 8:71–8.

120. Aparicio S, Caldas C. The implications of clonal genome evolution for cancer medicine. N Engl J Med 2013; 368:842–51.

121. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, Altman RB, Klein TE. Pharmacogenomics knowledge for personalized medicine. Clin Pharmacol Ther 2012; 92:414–7.

122. Griffith M, Griffith OL, Coffman AC, Weible JV, McMichael JF, Spies NC, Koval J, Das I, Callaway MB, Eldred JM, Miller CA, Subramanian J, Govindan R, et al. DGIdb: mining the druggable genome. Nat Methods 2013; 10:1209–10.

123. Kaplun A, Hogan JD, Schacherer F, Peter AP, Krishna S, Braun BR, Nambudiry R, Nitu MG, Mallelwar R, Albayrak A. PGMD: a comprehensive manually curated pharmacogenomic database. Pharmacogenomics J [Internet] 2015; Available from: http://dx.doi.org/10.1038/tpj.2015.32

124. Sosman JA, Kim KB, Schuchter L, Gonzalez R, Pavlick AC, Weber JS, McArthur GA, Hutson TE, Moschos SJ, Flaherty KT, Hersey P, Kefford R, Lawrence D, et al. Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. N Engl J Med 2012; 366:707–14.

125. Prahallad A, Sun C, Huang S, Di Nicolantonio F, Salazar R, Zecchin D, Beijersbergen RL, Bardelli A, Bernards R. Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. Nature 2012; 483:100–3.

126. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012; 486:346–52.

127. Fernandez-Retana J, Lasa-Gonsebatt F, Lopez-Urrutia E, Coronel-Martínez J, Cantu De Leon D, Jacobo-Herrera N, Peralta-Zaragoza O, Perez-Montiel D, Reynoso-Noveron N, Vazquez-Romo R, Perez-Plasencia C. Transcript profiling distinguishes complete treatment responders with locally advanced cervical cancer. Transl Oncol 2015; 8:77–84.

128. Gaba RC, Groth JV, Parvinian A, Guzman G, Casadaban LC. Gene expression in hepatocellular carcinoma: pilot

study of potential transarterial chemoembolization response biomarkers. J Vasc Interv Radiol 2015; 26:723–32.

129. Estevez-Garcia P, Rivera F, Molina-Pinelo S, Benavent M, Gómez J, Limón ML, Pastor MD, Martinez-Perez J, Paz-Ares L, Carnero A, Garcia-Carbonero R. Gene expression profile predictive of response to chemotherapy in metastatic colorectal cancer. Oncotarget 2015; 6:6151–9. doi: 10.18632/oncotarget.3152.

130. Costello JC, Heiser LM, Georgii E, Gönen M, Menden MP, Wang NJ, Bansal M, Ammad-ud-din M, Hintsanen P, Khan SA, Mpindi J-P, Kallioniemi O, Honkela A, et al. A community effort to assess and improve drug sensitivity prediction algorithms. Nat Biotechnol 2014; 32:1202–12.

131. Kao J, Salari K, Bocanegra M, Choi Y-L, Girard L, Gandhi J, Kwei KA, Hernandez-Boussard T, Wang P, Gazdar AF, Minna JD, Pollack JR. Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. PLoS One 2009; 4:e6146.

132. Geeleher P, Cox NJ, Huang RS. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. Genome Biol 2014; 15:R47.

133. Li B, Shin H, Gulbekyan G, Pustovalova O, Nikolsky Y, Hope A, Bessarabova M, Schu M, Kolpakova-Hart E, Merberg D, Dorner A, Trepicchio WL. Development of a Drug-Response Modeling Framework to Identify Cell Line Derived Translational Biomarkers That Can Predict Treatment Outcome to Erlotinib or Sorafenib. PLoS One 2015; 10:e0130700.

134. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, Bindal N, Beare D, Smith JA, Thompson IR, Ramaswamy S, Futreal PA, Haber DA, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res 2013; 41:D955–61.

135. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 2012; 483:603–7.

136. Reinhold WC, Sunshine M, Liu H, Varma S, Kohn KW, Morris J, Doroshow J, Pommier Y. CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. Cancer Res 2012; 72:3499–511.

137. Wilding JL, Bodmer WF. Cancer cell lines for drug discovery and development. Cancer Res 2014; 74:2377–84.

138. Sonkin D. Expression signature based on TP53 target genes doesn't predict response to TP53-MDM2 inhibitor in wild type TP53 tumors. Elife [Internet] 2015; 4. Available from: http://dx.doi.org/10.7554/eLife.10279

139. Haibe-Kains B, El-Hachem N, Birkbak NJ, Jin AC, Beck AH, Aerts HJWL, Quackenbush J. Inconsistency in large

pharmacogenomic studies. Nature 2013; 504:389–93.

140. Cancer Cell Line Encyclopedia Consortium, Genomics of Drug Sensitivity in Cancer Consortium. Pharmacogenomic agreement between two cancer cell line data sets. Nature 2015; 528:84–7.

141. Sonkin D, Hassan M, Murphy DJ, Tatarinova TV. Tumor suppressors status in cancer cell line Encyclopedia. Mol Oncol 2013; 7:791–8.

142. Ahmed D, Eide PW, Eilertsen IA, Danielsen SA, Eknæs M, Hektoen M, Lind GE, Lothe RA. Epigenetic and genetic features of 24 colon cancer cell lines. Oncogenesis 2013; 2:e71.

143. Singh Nanda J, Kumar R, Raghava GPS. dbEM: A database of epigenetic modifiers curated from cancerous and normal genomes. Sci Rep 2016; 6:19340.

144. Shen H, Laird PW. Interplay between the cancer genome and epigenome. Cell 2013; 153:38–55.

145. Easwaran H, Tsai H-C, Baylin SB. Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. Mol Cell 2014; 54:716–27.

146. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput Biol 2012; 8:e1002375.

147. Mutation Consequences and Pathway Analysis working group of the International Cancer Genome Consortium. Pathway and network analysis of cancer genomes. Nat Methods 2015; 12:615–21.

148. Kuperstein I, Grieco L, Cohen DPA, Thieffry D, Zinovyev A, Barillot E. The shortest path is not the one you know: application of biological network resources in precision oncology research. Mutagenesis 2015; 30:191–204.

149. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000; 25:25–9.

150. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics 2011; 27:1739–40.

151. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 1999; 27:29–34.

152. Nishimura D. BioCarta. Biotech Software & Internet Report 2001; 2:117–20.

153. Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, Caudy M, Garapati P, Gillespie M, Kamdar MR, Jassal B, Jupe S, Matthews L, et al. The Reactome pathway knowledgebase. Nucleic Acids Res 2014; 42:D472–7.

154. Kamburov A, Pentchev K, Galicka H, Wierling C, Lehrach H, Herwig R. ConsensusPathDB: toward a more complete picture of cell biology. Nucleic Acids Res 2011; 39:D712–7.

155. Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O, Anwar N, Schultz N, Bader GD, Sander C. Pathway

Commons, a web resource for biological pathway data. Nucleic Acids Res 2011; 39:D685–90.

156. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 2013; 41:D808–15.

157. Chatr-Aryamontri A, Breitkreutz B-J, Oughtred R, Boucher L, Heinicke S, Chen D, Stark C, Breitkreutz A, Kolas N, O'Donnell L, Reguly T, Nixon J, Ramage L, et al. The BioGRID interaction database: 2015 update. Nucleic Acids Res 2015; 43:D470–8.

158. Jewison T, Su Y, Disfany FM, Liang Y, Knox C, Maciejewski A, Poelzer J, Huynh J, Zhou Y, Arndt D, Djoumbou Y, Liu Y, Deng L, et al. SMPDB 2.0: big improvements to the Small Molecule Pathway Database. Nucleic Acids Res 2014; 42:D478–84.

159. Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, Buetow KH. PID: the Pathway Interaction Database. Nucleic Acids Res 2009; 37:D674–9.

160. An O, Pendino V, D'Antonio M, Ratti E, Gentilini M, Ciccarelli FD. NCG 4.0: the network of cancer genes in the era of massive mutational screenings of cancer genomes. Database 2014; 2014:bau015.

161. Kuperstein I, Bonnet E, Nguyen H-A, Cohen D, Viara E, Grieco L, Fourquet S, Calzone L, Russo C, Kondratova M, Dutreix M, Barillot E, Zinovyev A. Atlas of Cancer Signalling Network: a systems biology resource for integrative analysis of cancer data with Google Maps. Oncogenesis 2015; 4:e160.

162. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011; 144:646–74.

163. Villaveces JM, Jimenez RC, Habermann BH. KEGGViewer, a BioJS component to visualize KEGG Pathways. F1000Res 2014; 3:43.

164. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003; 13:2498–504.

165. Saito R, Smoot ME, Ono K, Ruscheinski J, Wang P-L, Lotia S, Pico AR, Bader GD, Ideker T. A travel guide to Cytoscape plugins. Nat Methods 2012; 9:1069–76.

166. Kutmon M, van Iersel MP, Bohler A, Kelder T, Nunes N, Pico AR, Evelo CT. PathVisio 3: an extendable pathway analysis toolbox. PLoS Comput Biol 2015; 11:e1004085.

167. Dixon P, Higginson I. AIDS and cancer pain treated with slow release morphine. Postgrad Med J 1991; 67 Suppl 2:S92–4.

168. MetaCore | Thomson Reuters [Internet]. [cited 2015 Oct 2]; Available from: http://www.thomsonreuters.com/metacore

169. Elsevier. Biological Research – Pathway Studio | Elsevier [Internet]. [cited 2015 Oct 2]; Available from: http://www.elsevier.com/solutions/pathway-studio

170. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 2003; 4:P3.

171. Wang J, Duncan D, Shi Z, Zhang B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res 2013; 41:W77–83.

172. Finotello F, Di Camillo B. Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis. Brief Funct Genomics 2015; 14:130–42.

173. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005; 102:15545–50.

174. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 2013; 14:7.

175. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C, Fröhling S, Chan EM, Sos ML, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature 2009; 462:108–12.

176. Tarca AL, Draghici S, Khatri P, Hassan SS, Mittal P, Kim J-S, Kim CJ, Kusanovic JP, Romero R. A novel signaling pathway impact analysis. Bioinformatics 2009; 25:75–82.

177. Haynes WA, Higdon R, Stanberry L, Collins D, Kolker E. Differential expression analysis for pathways. PLoS Comput Biol 2013; 9:e1002967.

178. Jia P, Zhao Z. Personalized pathway enrichment map of putative cancer genes from next generation sequencing data. PLoS One 2012; 7:e37595.

179. Oshlack A, Wakefield MJ. Transcript length bias in RNA-seq data confounds systems biology. Biol Direct 2009; 4:14.

180. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol 2010; 11:R14.

181. Wang X, Cairns MJ. SeqGSEA: a Bioconductor package for gene set enrichment analysis of RNA-Seq data integrating differential expression and splicing. Bioinformatics 2014; 30:1777–9.

182. Mi G, Di Y, Emerson S, Cumbie JS, Chang JH. Length bias correction in gene ontology enrichment analysis using logistic regression. PLoS One 2012; 7:e46128.

183. Xiong Q, Ancona N, Hauser ER, Mukherjee S, Furey TS. Integrating genetic and gene expression evidence into genome-wide association analysis of gene sets. Genome Res 2012; 22:386–97.

184. Mo Q, Wang S, Seshan VE, Olshen AB, Schultz N, Sander C, Powers RS, Ladanyi M, Shen R. Pattern discovery and cancer gene identification in integrated cancer genomic data. Proc Natl Acad Sci U S A 2013; 110:4245–50.

185. MacNeil SM, Johnson WE, Li DY, Piccolo SR, Bild AH. Inferring pathway dysregulation in cancers from multiple types of omic data. Genome Med 2015; 7:61.

186. Vaske CJ, Benz SC, Sanborn JZ, Earl D, Szeto C, Zhu J, Haussler D, Stuart JM. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. Bioinformatics 2010; 26:i237–45.

187. Ng S, Collisson EA, Sokolov A, Goldstein T, Gonzalez-Perez A, Lopez-Bigas N, Benz C, Haussler D, Stuart JM. PARADIGM-SHIFT predicts the function of mutations in multiple cancers using pathway impact analysis. Bioinformatics 2012; 28:i640–6.

188. Schroeder MP, Gonzalez-Perez A, Lopez-Bigas N. Visualizing multidimensional cancer genomics data. Genome Med 2013; 5:9.

189. Perez-Llamas C, Lopez-Bigas N. Gitools: analysis and visualisation of genomic data using interactive heat-maps. PLoS One 2011; 6:e19541.

190. Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Tamborero D, Schroeder MP, Jene-Sanz A, Santos A, Lopez-Bigas N. IntOGen-mutations identifies cancer drivers across tumor types. Nat Methods 2013; 10:1081–2.

191. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res 2009; 19:1639–45.

192. Medvedev P, Stanciu M, Brudno M. Computational methods for discovering structural variation with next-generation sequencing. Nat Methods 2009; 6:S13–20.

193. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 2012; 22:568–76.

194. Neuman JA, Isakov O, Shomron N. Analysis of insertion-deletion from deep-sequencing data: software evaluation for optimal detection. Brief Bioinform 2013; 14:46–55.

195. Ghoneim DH, Myers JR, Tuttle E, Paciorkowski AR. Comparison of insertion/deletion calling algorithms on human next-generation sequencing data. BMC Res Notes 2014; 7:864.

196. Wang Q, Jia P, Li F, Chen H, Ji H, Hucks D, Dahlman KB, Pao W, Zhao Z. Detecting somatic point mutations in cancer genome sequencing data: a comparison of mutation callers. Genome Med 2013; 5:91.

197. Wang C, Davila JI, Baheti S, Bhagwate AV, Wang X, Kocher J-PA, Slager SL, Feldman AL, Novak AJ, Cerhan JR, Thompson EA, Asmann YW. RVboost: RNA-seq variants prioritization using a boosting method. Bioinformatics 2014; 30:3414–6.

198. Piskol R, Ramaswami G, Li JB. Reliable identification of genomic variants from RNA-seq data. Am J Hum Genet 2013; 93:641–51.

199. Radenbaugh AJ, Ma S, Ewing A, Stuart JM, Collisson EA, Zhu J, Haussler D. RADIA: RNA and DNA integrated analysis for somatic mutation detection. PLoS One 2014; 9:e111516.

200. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, Janoueix-Lerosey I, Delattre O, Barillot E. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics 2012; 28:423–5.

201. Magi A, Tattini L, Cifola I, D'Aurizio R, Benelli M, Mangano E, Battaglia C, Bonora E, Kurg A, Seri M, Magini P, Giusti B, Romeo G, et al. EXCAVATOR: detecting copy number variants from whole-exome sequencing data. Genome Biol 2013; 14:R120.

202. Favero F, Joshi T, Marquard AM, Birkbak NJ, Krzystanek M, Li Q, Szallasi Z, Eklund AC. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. Ann Oncol 2015; 26:64–70.

203. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012; 487:330–7.

204. Hoffmann S, Otto C, Doose G, Tanzer A, Langenberger D, Christ S, Kunz M, Holdt LM, Teupser D, Hackermüller J, Stadler PF. A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. Genome Biol 2014; 15:R34.

205. Tang J, Fang F, Miller DF, Pilrose JM, Matei D, Huang TH-M, Nephew KP. Global DNA methylation profiling technologies and the ovarian cancer methylome. Methods Mol Biol 2015; 1238:653–75.

206. Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat Protoc 2013; 8:1765–86.

207. Feng H, Qin Z, Zhang X. Opportunities and methods for studying alternative splicing in cancer with RNA-Seq. Cancer Lett 2013; 340:179–91.

208. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013; 31:213–9.

209. Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 2012; 28:1811–7.

210. Tang X, Baheti S, Shameer K, Thompson KJ, Wills Q, Niu N, Holcomb IN, Boutet SC, Ramakrishnan R, Kachergus JM, Kocher J-PA, Weinshilboum RM, Wang L, et al. The eSNV-detect: a computational system to identify expressed single nucleotide variants from transcriptome sequencing data. Nucleic Acids Res 2014; 42:e172.

211. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel:

a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 2009; 25:2865–71.

212. Albers CA, Lunter G, MacArthur DG, McVean G, Ouwehand WH, Durbin R. Dindel: accurate indel calls from short-read data. Genome Res 2011; 21:961–73.

213. Narzisi G, O'Rawe JA, Iossifov I, Fang H, Lee Y-H, Wang Z, Wu Y, Lyon GJ, Wigler M, Schatz MC. Accurate de novo and transmitted indel detection in exome-capture data using microassembly. Nat Methods 2014; 11:1033–6.

214. Yoon S, Xuan Z, Makarov V, Ye K, Sebat J. Sensitive and accurate detection of copy number variants using read depth of coverage. Genome Res 2009; 19:1586–92.

215. Xie C, Tammi MT. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinformatics 2009; 10:80.

216. Boeva V, Zinovyev A, Bleakley K, Vert J-P, Janoueix-Lerosey I, Delattre O, Barillot E. Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization. Bioinformatics 2011; 27:268–9.

217. Xi R, Hadjipanayis AG, Luquette LJ, Kim T-M, Lee E, Zhang J, Johnson MD, Muzny DM, Wheeler DA, Gibbs RA, Kucherlapati R, Park PJ. Copy number variation detection in whole-genome sequencing data using the Bayesian information criterion. Proc Natl Acad Sci U S A 2011; 108:E1128–36.

218. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, Laird PW, Onofrio RC, Winckler W, Weir BA, Beroukhim R, Pellman D, Levine DA, et al. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol 2012; 30:413–21.

219. Louhimo R, Lepikhova T, Monni O, Hautaniemi S. Comparative analysis of algorithms for integration of copy number and expression data. Nat Methods 2012; 9:351–5.

220. Xing M. Molecular pathogenesis and mechanisms of thyroid cancer. Nat Rev Cancer 2013; 13:184–99.

221. Kim D, Salzberg SL. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biol 2011; 12:R72.

222. Jia W, Qiu K, He M, Song P, Zhou Q, Zhou F, Yu Y, Zhu D, Nickerson ML, Wan S, Liao X, Zhu X, Peng S, et al. SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. Genome Biol 2013; 14:R12.

223. Torres-García W, Zheng S, Sivachenko A, Vegesna R, Wang Q, Yao R, Berger MF, Weinstein JN, Getz G, Verhaak RGW. PRADA: pipeline for RNA sequencing data analysis. Bioinformatics 2014; 30:2224–6.

224. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, V Lord R, Clark SJ, Molloy PL. De novo identification of differentially methylated regions in the human genome. Epigenetics Chromatin 2015; 8:6.

225. Pedersen BS, Schwartz DA, Yang IV, Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. Bioinformatics 2012; 28:2986–8.

226. Li S, Garrett-Bakelman FE, Akalin A, Zumbo P, Levine R, To BL, Lewis ID, Brown AL, D'Andrea RJ, Melnick A, Mason CE. An optimized algorithm for detecting and annotating regional differential methylation. BMC Bioinformatics 2013; 14 Suppl 5:S10.

227. Witte T, Plass C, Gerhauser C. Pan-cancer patterns of DNA methylation. Genome Med 2014; 6:66.

228. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015; 43:e47.

229. González-Vallinas M, Vargas T, Moreno-Rubio J, Molina S, Herranz J, Cejas P, Burgos E, Aguayo C, Custodio A, Reglero G, Feliu J, Ramírez de Molina A. Clinical relevance of the differential expression of the glycosyltransferase gene GCNT3 in colon cancer. Eur J Cancer 2015; 51:1–8.

230. Shukla HD, Mahmood J, Vujaskovic Z. Integrated proteo-genomic approach for early diagnosis and prognosis of cancer. Cancer Lett [Internet] 2015; Available from: http://dx.doi.org/10.1016/j.canlet.2015.08.003

231. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. Genome Res 2012; 22:2008–17.

232. Wang W, Qin Z, Feng Z, Wang X, Zhang X. Identifying differentially spliced genes from two groups of RNA-seq samples. Gene 2013; 518:164–70.

233. Shen S, Park JW, Huang J, Dittmar KA, Lu Z-X, Zhou Q, Carstens RP, Xing Y. MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. Nucleic Acids Res 2012; 40:e61.

234. Lend AK, Kazantseva A, Kivil A, Valvere V, Palm K. Diagnostic significance of alternative splice variants of REST and DOPEY1 in the peripheral blood of patients with breast cancer. Tumour Biol 2015; 36:2473–80.

235. Jackson C, Browell D, Gautrey H, Tyson-Capper A. Clinical Significance of HER-2 Splice Variants in Breast Cancer Progression and Drug Resistance. Int J Cell Biol 2013; 2013:973584.