11

Bioinformatics in Cancer Detection

Mohit Sharma[1,2], Umme Abiha[3], Parul Chugh[3], Balakumar Chandrasekaran[4] and Nikhil Agrawal[5]*

[1]Post Graduate School of Molecular Medicine, Medical University of Warsaw, Warsaw, Poland
[2]Malopolska Centre of Biotechnology, Jagiellonian University, Krakow, Poland
[3]Amity Institute of Biotechnology, Noida, India
[4]Faculty of Pharmacy, Philadelphia University-Jordan, Amman, Jordan
[5]College of Health Sciences, University of KwaZulu-Natal, Westville, Durban, South Africa

Abstract

The uncontrolled growth of abnormal cells is generally determined by the alterations in genes and epigenetic factors. Oncological research is undergoing a drastic revolution due to advanced technology improvement in the verge of exploring the relationship of molecules which make a unit cell of an organism. Molecular biologists are more aware about the genomic, transcriptomic, and proteomic data because of the obvious technology advancement in the area of molecular medicine. This has further lead to the development of novel potential targets for drug development and also establishment of molecular markers for unified treatment and therapy against cancer. Numerous cancer studies have been carried out using altered protocols, samples, and data from multiple sources in order to compare and validate new strategies with the conventional ones. Moreover, it also opens a wide arena to develop personalized or stratified medicine to counter medicinal upheaval. Bioinformatics helps to develop new methods and advancing trends in order to attain the ultimate goal of developing therapeutics and diagnostic protocols in the area of cancer research.

In this chapter, we will discuss about the contributions, applications, and importance of bioinformatics in cancer research.

**Corresponding author*: nikhil.08oct@gmail.com

S. Balamurugan, Anand Krishnan, Dinesh Goyal, Balakumar Chandrasekaran and Boomi Pandi (eds.) Computation in Bioinformatics: Multidisciplinary Applications, (229–244)

***Keywords*:** Cancer, bioinformatics, epigenetic factors

11.1 Introduction

Cancer is the most deadly and devastating disease, prevalence of which has claimed millions of lives worldwide. Despite advanced research and technologies that have evolved in the area of clinical diagnosis and precision medicine, cancer still remains a dreaded disease with serious health implications. This disease can either be a cause of genetic or epigenetic factors or even both. WHO reports the burden of cancer to be claiming 9.6 million lives in an estimated annual statistics record of 2018. It has been reported to be a cause of death among 1 in 6 deceased worldwide and remains an immense medical challenge [1]. Various genetic alterations and molecular aberrations aggravate cancer into a progressive disease; therefore, the best possible therapeutic intervention must target the alterations and abnormalities to evade the proliferative growth. Genetic abnormalities such as mutation, translocation, deletion, replication, and even post-translational modifications often lead to different types of cancer. Therefore, before inducing or initiating a therapy to any cancer patient, the etiological origin should be known and the exact aberrations and alteration patterns of a particular type of cancer should be confirmed. Testing through drug sensitivity, cancer bioinformatics and pharmacogenetics are major parts of PCT that are designed to unravel the genetic alterations and molecular abnormality information and select optimal anticancer drugs [2].

Use of advanced computing technology and mathematical approach to store, manage, and analyze data is one the foremost ground on which bioinformatics take a stakehold in the field of P-4 medicine. The biggest hindrance that cancer researchers faces is the lack of infrastructure to store and analyze biological data, but with the use of virtual repositories, this has improved the problem as it is readily available in the public domain from all institutions [3].

11.2 The Era of Bioinformatics in Cancer

The field of bioinformatics in the area of cancer prevention and treatment has swiftly progressed over the past many years. Since the onset of its initial implementation, epigenetic studies are now possible due to the advancement of high throughput techniques like next-generation sequencing, ChiP-ChiP, ChiP-Seq, and high-density microarrays for miRNA profiling

to detect the dysregulation of tumor suppressor genes. This has further altered our understanding of cancer biology, leading to progressively competitive approaches to investigate additional and advanced datasets [4]. Bioinformatic pipelines work in conjunction of predicting a massive influence of genetic aberrations and quantifying changes of the tumor microenvironment. Extensive research has led to the development of new tools that analyze tumor-immune microenvironment and its interactions to assess tumor infiltrating leukocyte content, microsatellite instability, total alteration burden, and neoantigen presentation [5]. The enormous growth of information and technology in the area of genomics and omics is flabbergasting, as massive amounts of data is being generated with sequencing and microarray chips to integrate omics-based analysis. Furthermore, this has accelerated machine learning and integrated network–based approaches on multiple platforms to investigate assorted data within the reach of public accessible resources, like the cancer order atlas [6].

Annotating molecular and genetic aberrations as infiltrating cell states in various ideal sequencing conditions and application specific approaches are on improving verges to establish strong analysis techniques and measuring gene expression. This ability provides a scientific research to elucidate the mechanism that holds the process of genetic expression and its upheaval. These techniques that vary from improved quantification of copy variety and organic phenomenon from formal mounted tissues as applications that need high sensitivity like the quantification of tumor mutations from liquid biopsies (circulating cell free DNA) [7]. As a rigorous need to develop an efficient knowledge of understanding the constraints of processing approaches along with novel techniques to improve the flexibility of distribution and molecular impact of genes like repetitive or permutable endogenous parts, an increasing scope is felt to build well-annotated databases of simply accessible information in multi-variable

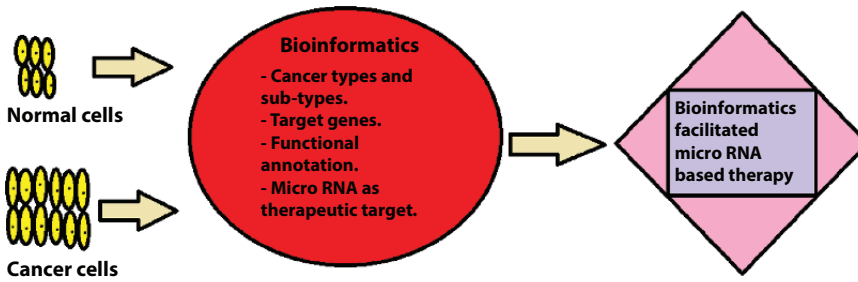

Figure 11.1 Role of bioinformatics in cancer detection.

analysis pipelines. With the advancement and isolation of SIGdb and cBioPortal, it can establish new diagnostic and prognostic biomarkers for traditional interventional modalities still as rising areas like immuno-oncology and areas of unmet clinical want [8]. This highlights the present state of bioinformatics applications in cancer biology and infers future prospects for rising information processing applications through computer science and machine learning approaches. Figure 11.1 explains the role of bioinformatics in cancer research through facilitated micro-RNA–based therapy.

11.3 Aid in Cancer Research via NCI

National Cancer Institute or NCI has been instrumental in the advancement of cancer research by evolving the field of genomics, proteomics, and metabolomics to integrate the datasets and develop the know-how of etiological origin and molecular basis of cancers [9]. The Center for Biomedical Informatics and Information Technology (CBIIT) supervises computational data and exchange of bioinformatics-computed results among various institutions, where National Cancer Informatics Program (NCIP) uses NCIP HubExit Disclaimer and integrates the genomic, clinical, and translational studies to enhance sharing of records, regular assessments, and data visualization. It is a central operating platform, designed for researchers to exchange, share, and record important bioinformatics data to speed up cancer research. The operating platform of NCIP is a test, whether the cancer research community reveals the social and network factors of this machine beneficial for crew technological expertise and multi-investigator study teams or not [10]. NCI is a public repository that guarantees saving clinical data for a long time with collaborations among private companies to foster technology advancement and come up with fueled diagnosis and treatment approaches. This is achieved in consolidation to the Cancer Data Science Laboratory (CDSL), in NCI's Center for Cancer Research which primarily functions to collaborate laboratory statistics from cancer genomics and omics signatures and further develop computational technology to investigate it. Together, computational simulations and mathematical algorithms are instrumental in answering the critical research questions, like the etiological origin of cancer, its progression, and how can it be prevented or cured.

The Cancer Genome Atlas (TCGA) was a collaborative three year pilot project, initiated by NCI and National Human Genome Research Institute in 2005 to report various mutations in human genome that are associated

with cancer progression. Researchers used high throughput genome analysis technique on more than 11,000 cancer patients with tumor and normal-healthy samples, on which various researches have been conducted and published [11]. NCI's Therapeutically Applicable Research to Generate Effective Treatments (TARGET) program has enabled scientists to isolate genetic aberrations in pediatric cancers which is predominantly observed in children during medical trials carried out by Children's Oncology Group. NCI's Clinical Proteomic Tumor Analysis Consortium (CPTAC) is a constitutive national collaboration of institutions and researchers that are involved in genomic and proteomic analysis of cancer to develop a ground of understanding on the prevalence of cancer. NCI coordinates clinical trials and provides an extensive support to intramural researchers and further foster biomedical research which is also available in public domain [12]. This extensive network has fueled the understanding of biology of cancer and isolated potential targets for novel drug-targeted therapy. Invariably, this has surfaced the establishment and development of mounted statistics and been able to discover new equipments and approaches to analyze and access the big data.

11.4 Application of Big Data in Developing Precision Medicine

The availability of volumes of biological data in the public domain was only possible because of the powerful new research technologies that evolved with time. The data has not only made the scientific community excited about the precision medicine but has also enabled the scientists and researchers to know the biological reason behind the disease.

This has made possible interventions with tailored medicines, evolution of preventive techniques, and diagnostic treatments.

The volumes of big data that are available in the databases can be retrieved by the researchers and exchanged on a platform to be useful in order to search for biological conclusions for the rare and unsolved mysteries of these deadly diseases [13].

The biggest limitation of cancer data is the fragmentation and compartmentalization; therefore, the researchers are trying to overcome the challenges that pose for fostering research. In order to speed the research, it is pertinent for the researchers to have an exchange and access of the curated data at the same time so that they are able to analyze the problem [14]. Therefore, there is a need of an infrastructure where researchers can store, analyze, visualize, integrate, and access the biological data on a common

public platform which remains accessible to all the researchers and scientist all over the globe. Bioinformatics is an evolving field of study that has been observing a boom past a decade. It uses advanced computing and coding algorithms, mathematic simulations, and different technological platforms for managing, storing, and analyzing the data. Nowadays, researchers are using different tools and platforms to manage the biological data from a comprehensive analysis of the proteins, whole genome sequencing, and advanced imaging studies but integrating the data from various platforms poses a difficulty as the researchers are unable to have the access of the primary data which was created by other studies and also due to the lack of computational tools and infrastructure necessary to integrate and analyze it [15]. Recently, a big splurged boom in the use of virtual repositories, also known as Data clouds, has not only helped to integrate but also improve

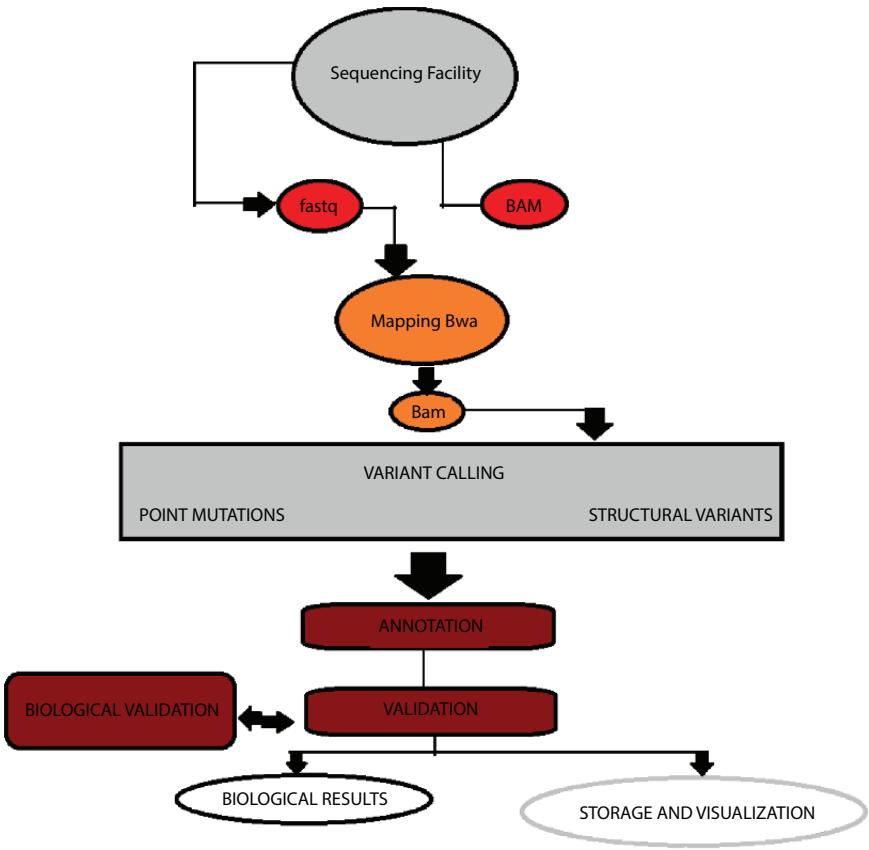

Figure 11.2 Pipeline analysis.

the access to research data. Although these steps still remain in the fetal stages and questions surround the integrated organization and coordination of these clouds and their uses. It has been possible to understand the molecular basis of cancer because of NCI that played a leading role in scientific advancement of proteomics, genomics, and metabolomics [16]. The National Cancer Informatics Program is playing a pivotal role in including clinical, genomic, and translational studies to improve the data sharing analysis and visualization in numerous research areas [17, 18]. For example, Figure 11.2 gives a detailed pipeline analysis of the structural variants and point mutations of a cancer and normal cell that can be reported via variable sequencing facilities where the data is annotated and integrated. The data is further compiled for undergoing either validation through wet laboratory procedures or a course of discussion in scientific community. The validation of the data develops a biological ground of systemic understanding which not only leads to reliable biological results but can also be stored and visualized [19].

11.5 Historical Perspective and Development

Bioinformatics and computational revolution has strengthened the approach of understanding the genetic and molecular basis of cancer. This has led the scientific community to study significant gene expressions of the expressed genome, instead of looking at the individual genes. The abundance of molecular knowledge produced, both from the laboratory as well as the amount of data stored in the patient record keeps on growing at an incredible pace. Creating fresh perspectives into the cancer genetics has been critical in discovering new ways to incorporate these results. Bioinformatics, which is the integration of genetics, information technology, and engineering, continues to evolve as a key component of research into cancer biology [20].

Cancer bioinformatics is one of several ways of focusing bioinformatics approaches in cancer, depending on the nature of disease metabolic rates, signaling, interaction, and multiplication. Clinical bioinformatics, an evolving science incorporating clinical computer science, bioinformatics, information technology, medical informatics, mathematics, and omics research, can be regarded as one of the key aspects resolving clinically important problems in early detection, successful treatments, and reliable diagnosis of cancer patients. The development of cancer-specific bioinformatics methodologies or the implementation of new and specialized bioinformatics tools to respond to the specific cancer question is desired [21].

For instance, the Semantic Web platform has been used to interpret high-throughput clinical data and to create quantitative semantic models obtained from Corvus, a data warehouse that offers a standardized interface to different types of Omics data, focused on structured biological information while using SPARQL endpoint. Semantic models including transcriptomic, epigenomic, and genomic data from melanoma samples with data from Gene Ontology and regulatory networks built from transcription factor binding information were used for the interplay between a cell molecular state and its reaction to anticancer therapy. Multivariate assays, a method for characterizing the error produced in the assay outcomes by the intrinsic error in the preparation of samples and the calculation of contributing variables, are used to support and direct clinicians to consider the use of PAM50 centroid-based genomic predictors for breast cancer care strategies and to include useful knowledge regarding the ambiguity. It can be taken seriously into account the applicability, accuracy, and convergence of methodologies, applications, computing methods, and databases that can be used to investigate cancer molecular pathways and recognize and verify novel biomarkers, network biomarkers, and individualized cancer medicines.

miRTrail is an integrative method for evaluating detailed gene and miRNA associations dependent on expression profiles to produce more rigorous and accurate data on pathogenic deregulation processes. It was proposed that miRTrail could open up opportunities to explore regulatory correlations between genes and miRNAs for human diseases, including cancer, by combining knowledge on 20,000 genes, approximately 1,000 miRNAs, and around 280,000 putative interactions. Exploring the theoretical computational mode that compares certain regulatory interactions between genes and miRNAs with clinical phenotypes would be useful, e.g., variance in gene interactions across tumor sites, phages, patient effects, or therapy responses [22].

Medical imaging must be one of the key considerations to be addressed in the development of cancer bioinformatics, as imaging in clinical pathology, tomography, NMR imaging, and positron emission tomography are one of the most appropriate and effective methods in the identification and diagnosis of cancer "early and precise". Bioinformatics-based evaluations of surface morphology of masses and other anomalies in medical images are undertaken by systematic extraction of target characteristics by mathematical morphology and by two contrast adjustment techniques improving the extracted characteristics. Based on clinical breast cancer data [23], the algorithm mentioned by Haustein and Schumacher in the Thematic Series on Cancer Bioinformatics in Clinical Bioinformatics Journal can predict

tumor growth and determine the development of some metastases prior to clinical detection in cells. It could be a non-relative issue or a possible hope if cancer bioinformatics specialists will assist clinicians in defining the possible image of gene or protein associations and pathways aligned with tumor-associated shapes, densities, or positions.

11.6 Bioinformatics-Based Approaches in the Study of Cancer

There are many bioinformatics approaches to study cancers that categorize different types of cancers either on the basis of gene expression or genetic profiles. Figure 11.3 depicts three instrumental approaches: stepwise linkage analysis of microarray signatures (SLAMS), module maps, and cancer outlier profile analysis (COPA).

11.6.1 SLAMS

In order to signify the standard of care for cancer, tumors are categorized on the basis of grade, stage, and histology. Detailed analysis of histological specimens gives just a view at the molecular level, and therefore, various types of tissue appear identical but do not behave the same and several associations are not characteristic. These molecular variations are probably responsible for the result variation and treatment response for closely categorized tissues. This adds to substantial confusion as doctors have to precisely adapt a particular treatment for the patient [24]. While profiles

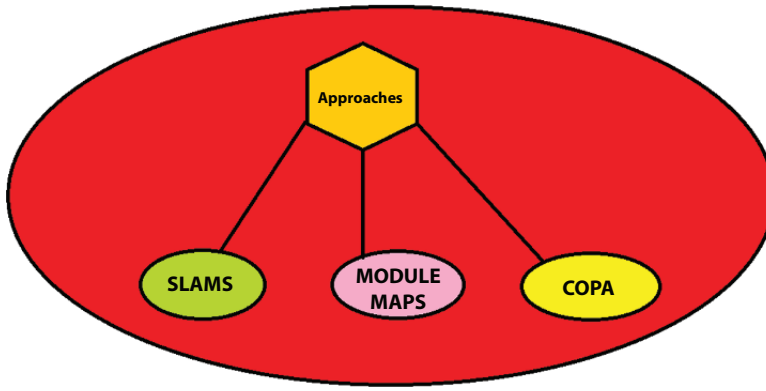

Figure 11.3 Different bioinformatics-based approaches.

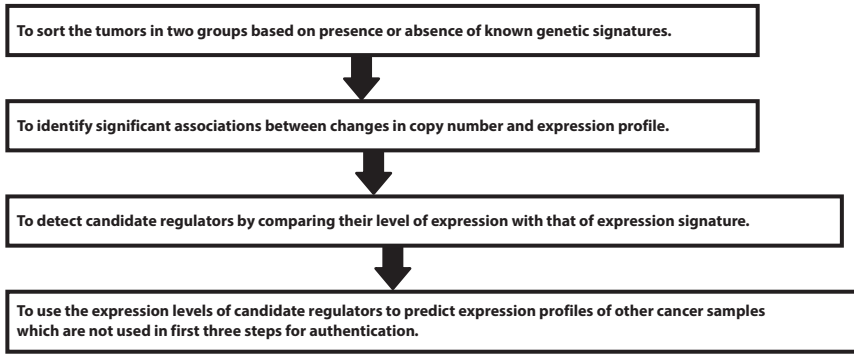

Figure 11.4 Steps involved in SLAMS.

of gene expression are not widely utilized in clinical treatment, it provides confirmation of a more systematic approach to tumor detection. Even so, this method still has its drawbacks, because the end result is not only to discover genetic markers but rather to recognize how the genes work in the disease and to establish potential targets for impacting output.

A group of scientists at Stanford University has developed a method for identifying data on gene expression by shifting the number of copies of DNA to chromosome regions consisting of candidate oncogene regulators. The gene expression profile associated with a known tumor type [25–27] is the phenotype in the linkage study.

Their method, i.e., SLAMS consists of four steps (Figure 11.4).

SLAMS was confirmed by utilizing it on the breast cancer samples and was subsequently concluded that a gene in proximity might interact with the MYC gene (oncogenic transcription factor) to express wound signature [28].

11.6.2 Module Maps

Module maps may be a possible tool [29, 30] for illustrating the normal trends of gene expression through heterogeneous tissues and cancer disease processes. Modules can be calculated by comparing specific gene sets with expression data and by retrieving a subset of co-expressed genes. The modules so acquired can be applied to all types of tissues to look for common signatures, exposing underlying processes.

Article published in Nature Genetics [31], by using module maps, indicates the expectations for cancer studies. The scientists obtained findings for various types of cancers from 2,000 microarray research and compiled

them with clinical evidence. About 456 modules were identified from 3,000 gene sets and then matched across multiple cancers. The study found variations in gene expression common to different cancers, thereby offering valuable knowledge about pathways for cancer. This technique is helpful in highlighting dynamics of global disease that are not apparent from studies with a single microarray. But there are also limits. The inconsistency in outcomes from numerous techniques presents challenges as well as the lack of accurate clinical knowledge for analysis trials will influence the utility [24, 32].

11.6.3 COPA

Contemporary oncogene expression profile identification techniques are limited in their ability to distinguish patterns through multiple samples, particularly when those patterns do not predominate enough to stand out. This is attributed to the heterogeneity of patterns of expression in samples which makes it difficult to differentiate from background variations between original patterns [33]. COPA is used to address this issue to classify the variants that are found in a subset of tumor samples. From this approach, we can distinguish the outer expression profiles for the specific group of genes that have lower levels of expression, and overexpression reveals a small subset of tumor samples. COPA technique comprises of three steps (Figure 11.5).

COPA facilitates in compacting normal dynamic range of the expression profiles. It also illuminates profiles that deviate from the mean expression level. For each gene, the degree of divergence from the normal pattern of expression is then rated to produce a suitable list of outer genes.

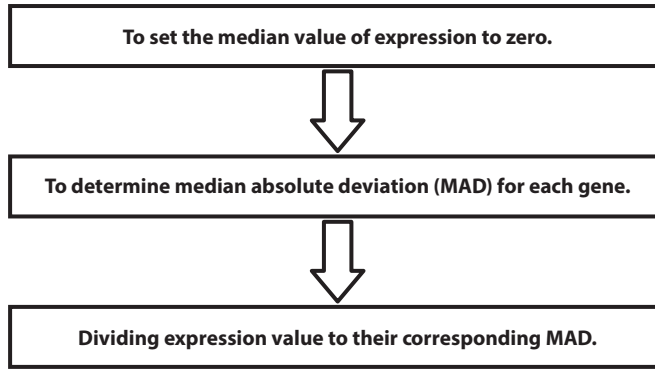

Figure 11.5 Steps involved in COPA.

COPA has been added to the oncomine database, defining a variety of outlier profiles [34]. Advances in COPA are prompting researchers to look at other tumors that may display similar rearrangements which may offer hints to potential diagnostic methods or therapeutic targets.

11.7 Conclusion and Future Challenges

There are multiple applications of bioinformatics in cancer research, but on the other hand, substantial challenges exist. More innovative approaches and techniques are needed to understand the poorly understood genetic changes associated with cancer. Methodology is also necessary for data integration and normalization algorithms for samples assembled under various conditions in different labs. Standard formats should be adopted for storing data and annotating samples.

Research is needed as positive or negative for cancer beyond the microarray sample mark; more thorough annotation is needed to understand the function of genetics in cancer subtypes, response to therapy, and other related parameters. Also, linking of microarray data with patient's clinical information should be done in computable mode and with storage of data and comes the challenge of effective mining of clinical data. caBIG™ is one of the biggest prospective to change the standard research procedure.

References

1. Ma, X. and Yu, H., Global burden of cancer. *Yale J. Biol. Med.*, 79, 3–4, 85–94, 2006.
2. Hasty, P. and Montagna, C., Chromosomal Rearrangements in Cancer: Detection and potential causal mechanisms. *Mol. Cell. Oncol.*, 1, 1, e29904, 2014, https://doi.org/10.4161/mco.29904.
3. Institute of Medicine (US) and National Academy of Engineering (US) Roundtable on Value & Science-Driven Health Care, *Engineering a Learning Healthcare System: A Look at the Future: Workshop Summary*, p. 3, National Academies Press (US), Washington (DC), 2011, Healthcare System Complexities, Impediments, and Failures. Available from: https://www.ncbi.nlm.nih.gov/books/NBK61963/.
4. Shyr, D. and Liu, Q., Next generation sequencing in cancer research and clinical application. *Biol. Proced. Online*, 15, 1, 4, 2013, https://doi.org/10.1186/1480-9222-15-4.

5. Hackl, H., Charoentong, P., Finotello, F. *et al.*, Computational genomics tools for dissecting tumour–immune cell interactions. *Nat. Rev. Genet.*, 17, 441–458, 2016, https://doi.org/10.1038/nrg.2016.67.
6. Misra, B., Langefeld, C., Olivier, M., Cox, L., Integrated omics: tools, advances and future approaches. *J. Mol. Endocrinol.*, 62, 1, R21–R45, 2019.
7. Elazezy, M. and Joosse, S.A., Techniques of using circulating tumor DNA as a liquid biopsy component in cancer management. *Comput. Struct. Biotechnol. J.*, 16, 370–378, 2018, https://doi.org/10.1016/j.csbj.2018.10.002.
8. Cesano, A. and Warren, S., Bringing the next Generation of Immuno-Oncology Biomarkers to the Clinic. *Biomedicines*, 6, 1, 14, 2018, https://doi.org/10.3390/biomedicines6010014.
9. Manzoni, C., Kia, D.A., Vandrovcova, J., Hardy, J., Wood, N.W., Lewis, P.A., Ferrari, R., Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. *Briefings Bioinf.*, 19, 2, 286–302, 2018, https://doi.org/10.1093/bib/bbw114.
10. Mongkolwat, P., Kleper, V., Talbot, S., Rubin, D., The National Cancer Informatics Program (NCIP) Annotation and Image Markup (AIM) Foundation model. *J. Digit. Imaging*, 27, 6, 692–701, 2014, https://doi.org/10.1007/s10278-014-9710-3.
11. Tomczak, K., Czerwińska, P., Wiznerowicz, M., The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp. Oncol. (Poznan, Poland)*, 19, 1A, A68–A77, 2015, https://doi.org/10.5114/wo.2014.47136.
12. Rudnick, P.A., Markey, S.P., Roth, J., Mirokhin, Y., Yan, X., Tchekhovskoi, D.V., Edwards, N.J., Thangudu, R.R., Ketchum, K.A., Kinsinger, C.R., Mesri, M., Rodriguez, H., Stein, S.E., A Description of the Clinical Proteomic Tumor Analysis Consortium (CPTAC) Common Data Analysis Pipeline. *J. Proteome Res.*, 15, 3, 1023–1032, 2016, https://doi.org/10.1021/acs.jproteome.5b01091.
13. Wooden, B., Goossens, N., Hoshida, Y., Friedman, S.L., Using Big Data to Discover Diagnostics and Therapeutics for Gastrointestinal and Liver Diseases. *Gastroenterology*, 152, 1, 53–67.e3, 2017, https://doi.org/10.1053/j.gastro.2016.09.065.
14. Tremblay, D., Touati, N., Roberge, D. *et al.*, Understanding cancer networks better to implement them more effectively: a mixed methods multicase study. *Implementation Sci.*, 11, 39, 2015, https://doi.org/10.1186/s13012-016-0404-8.
15. He, K.Y., Ge, D., He, M.M., Big Data Analytics for Genomic Medicine. *Int. J. Mol. Sci.*, 18, 2, 412, 2017, https://doi.org/10.3390/ijms18020412.
16. Chakraborty, S., Hosen, M., II, Ahmed, M., Shekhar, H.U., Onco-Multi-OMICS Approach: A New Frontier in Cancer Research. *BioMed. Res. Int.*, 2018, Article ID 9836256, https://doi.org/10.1155/2018/9836256.
17. National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Board on Health Care Services, National Cancer Policy Forum. *Improving Cancer Diagnosis and Care: Clinical Application of*

Computational Methods in Precision Oncology: Proceedings of a Workshop, 2019 Jun 20 National Academies Press (US), Washington (DC), Available from: https://www.ncbi.nlm.nih.gov/books/NBK544637/.

18. Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., Omberg, L., Wolf, D.M., Shriver, C.D., Thorsson, V., Cancer Genome Atlas Research Network, Hu, H., An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell*, 173, 2, 400–416.e11, 2018, https://doi.org/10.1016/j.cell.2018.02.052.
19. Rodriguez, H. and Pennington, S.R., Revolutionizing Precision Oncology through Collaborative Proteogenomics and Data Sharing. *Cell*, 173, 3, 535–539, 2018, https://doi.org/10.1016/j.cell.2018.04.008.
20. Hulsen, T., Jamuar, S.S., Moody, A.R., Karnes, J.H., Varga, O., Hedensted, S., Spreafico, R., Hafler, D.A., McKinney, E.F., From Big Data to Precision Medicine. *Front. Med.*, 6, 34, 2019, https://doi.org/10.3389/fmed.2019.00034.
21. Wu, D., Rice, C.M., Wang, X., Cancer bioinformatics: A new approach to systems clinical medicine. *BMC Bioinf.*, 13, 71, 2012, https://doi.org/10.1186/1471-2105-13-71.
22. Laczny, C., Leidinger, P., Haas, J. *et al.*, miRTrail - a comprehensive webserver for analyzing gene and miRNA patterns to enhance the understanding of regulatory mechanisms in diseases. *BMC Bioinf.*, 13, 36, 2012, https://doi.org/10.1186/1471-2105-13-36.
23. Wu, D., Rice, C., Wang, X., Cancer bioinformatics: A new approach to systems clinical medicine. *BMC Bioinf.*, 13, 71, 2012, 10.1186/1471-2105-13-71.
24. Park, T., Yi, S.G., Kang, S.H., Lee, S., Lee, Y.S., Simon, R., Evaluation of normalization methods for microarray data. *BMC Bioinf.*, 4, 33, 2003.
25. Chen, C.N., Lin, J.J., Chen, J.J., Lee, P.H., Yang, C.Y., Kuo, M.L., Chang, K.J., Hsieh, F.J., Gene expression profile predicts patient survival of gastric cancer after surgical resection. *J. Clin. Oncol.*, 23, 7286–95, 2005.
26. Dyrskjot, L., Zieger, K., Kruhoffer, M., Thykjaer, T., Jensen, J.L., Primdahl, H., Aziz, N., Marcussen, N., Moller, K., Orntoft, T.F., A molecular signature in superficial bladder carcinoma predicts clinical outcome, *Clin. Cancer Res.*, 11, 4029–36, 2005.
27. Gyorffy, B., Surowiak, P., Kiesslich, O., Denkert, C., Schafer, R., Dietel, M., Lage, H., Gene expression profiling of 30 cancer cell lines predicts resistance towards 11 anticancer drugs at clinically achieved concentrations, *Int. J. Cancer*, 118, 1699–712, 2006.
28. Gyorffy, B., Serra, V., Jurchott, K., Abdul-Ghani, R., Garber, M., Stein, U., Petersen, I., Lage, H., Dietel, M., Schafer, R., Prediction of doxorubicin sensitivity in breast tumors based on gene expression profiles of drug-resistant cell lines correlates with patient survival, *Oncogene*, 24, 51, 7542–51, 2005 Nov 17.

29. Ihmels, J., Friedlander, G., Bergmann, S., Sarig, O., Ziv, Y., Barkai, N., Revealing modular organization in the yeast transcriptional network, *Nat. Genet.*, 31, 370–7, 2002.
30. Tanay, A., Sharan, R., Kupiec, M., Shamir, R., Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data, *Proc. Natl. Acad. Sci. U.S.A.*, 101, 2981–6, 2004.
31. Segal, E., Friedman, N., Koller, D., Regev, A., A module map showing conditional activity of expression modules in cancer, *Nat. Genet.*, 36, 1090–8, 2004.
32. Chua, S.W., Vijayakumar, P., Nissom, P.M., Yam, C.Y., Wong, V.V., Yang, H., A novel normalization method for effective removal of systematic variation in microarray data, *Nucleic Acids Res.*, 34, e38, 2006.
33. Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., Lee, C., Montie, J.E., Shah, R.B., Pienta, K.J., Rubin, M.A., Chinnaiyan, A.M., Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer, *Science*, 310, 644–8, 2005.
34. Rhodes, D.R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pandey, A., Chinnaiyan, A.M., ONCOMINE: a cancer microarray database and integrated data-mining platform, *Neoplasia*, 6, 1–6, 2004.