



Opinion

# The High-Throughput Analyses Era: Are We Ready for the Data Struggle?

Valeria D'Argenio <sup>1,2</sup>

<sup>1</sup> CEINGE-Biotecnologie Avanzate, via G. Salvatore 486, 80145 Naples, Italy; dargenio@ceinge.unina.it; Tel.: +39-081-3737909

<sup>2</sup> Department of Molecular Medicine and Medical Biotechnologies, University of Naples Federico II, via Pansini 5, 80131 Naples, Italy

Received: 28 December 2017; Accepted: 27 February 2018; Published: 2 March 2018

**Abstract:** Recent and rapid technological advances in molecular sciences have dramatically increased the ability to carry out high-throughput studies characterized by big data production. This, in turn, led to the consequent negative effect of highlighting the presence of a gap between data yield and their analysis. Indeed, big data management is becoming an increasingly important aspect of many fields of molecular research including the study of human diseases. Now, the challenge is to identify, within the huge amount of data obtained, that which is of clinical relevance. In this context, issues related to data interpretation, sharing and storage need to be assessed and standardized. Once this is achieved, the integration of data from different -omic approaches will improve the diagnosis, monitoring and therapy of diseases by allowing the identification of novel, potentially actionably biomarkers in view of personalized medicine.

**Keywords:** high-throughput analysis; next-generation sequencing; big data; -omic sciences; personalized medicine.

## 1. Introduction

Rapid technological advances have made high-throughput technologies available for the study of biological systems in their integrity, laying the foundation for the development of the so-called “-omics” era [1,2]. Indeed, the completion of the first human genome sequence draft [3,4] and the availability of high-scale technological tools have made it possible to study genomics, transcriptomics, epigenomics, and other -omic sciences at a previously unthinkable level [2,5–7]. The integration of these disciplines is increasing our understanding of the molecular bases of human diseases (both acquired and inherited), with the final aim to improve their diagnosis, monitoring, and treatment, in view of an even more personalized medicine [2]. The currently available technologies can generate gigabytes of data per day with a good level of accuracy and reliability [2,5–7]. This feature has pushed molecular research beyond the limitations imposed by the more traditional analytical approaches. However, it was soon clear that the same feature badly conceals an important side effect: high-throughput technologies generate a high quantity of data, whose management, analysis and storage require specific infrastructures and bioinformatic pipelines [8,9]. In particular, the correct interpretation of these data to extrapolate, from the huge amount of information, only that which is relevant from a clinical point of view, represents nowadays a great challenge [10]. Also, ethical issues related to incidental findings, data property, and privacy aspects are animating scientific debates and need to be carefully regulated to avoid risks related to the data struggle.

## 2. High-Throughput Analyses

The last 15 years have featured the development and the fast diffusion of next-generation sequencing (NGS) technologies [2,5–7]. These techniques have impacted every field of molecular

research, escalating previously used sequencing technologies [11], and opening the way to the -omic sciences foundation [1,2]. Indeed, NGS methods allow the sequencing of entire genomes [12–15], of exomes [16–18], of panels of genes related to a disease of interest [19–21], or of a single gene [22–26], but can also be used to explore the entire transcriptome [27–29], small RNAs [30–32], the epigenome [33,34], and the microbiome [35–38].

Notwithstanding some peculiar characteristics related to the different manufacturers, the currently available NGS sequencers are based on the amplification of one specific library (or multiple barcoded libraries), i.e., a pool of DNA fragments representing the target to be sequenced, on the surface of a flow cell, or of specific microscopic beads, to obtain clonal clusters of fragments that will be massively sequenced, as is extensively reviewed elsewhere [5–7]. Next-generation sequencing techniques combine the high-throughput capability and a high sequencing reads accuracy with a low cost per base: it is amazing to think that the cost for the sequencing of an entire human genome has dropped from about US \$10 million to about US \$1000 in just the last 10 years [39].

In this view, it is not surprising that NGS is also becoming a reference method for molecular diagnostics [10]. In particular, NGS allows not only the analysis, in more patients simultaneously, of disease-related genes in less time and at lower costs than traditional approaches, but also the sequencing of panels of genes up to the complete exome [16–26]. In this way, it is possible to increase the diagnostic sensitivity, to discover novel disease-related genes and also obtain data regarding other genes that were potentially acting as disease-phenotype modifiers [19,40–42]. Due to their high sensitivity and flexibility, NGS methods are also useful for prenatal and preimplantation diagnostics [43–46], and other applications, such as the sequencing of circulating free DNA or of single cells (i.e., fetal cells or circulating tumor cells), are being validated [45,47–49].

In addition to the study of sequence variations at the DNA level, NGS methods can be used to study genetic variability, and the mechanisms underlying the onset of specific diseases at epigenetic, transcriptomic and metagenomic levels [27–38]. Indeed, several factors, other than individual genetic predisposition, such as diet, environmental factors and lifestyle, can influence the epigenome, the transcriptome and the microbiome [2,50]. Thus, all these systems are dynamic and can feature specific modifications related to a specific pathological status. Understanding such modifications not only sheds light on the mechanisms underlying the disease development, but may also provide novel, potential biomarkers for an earlier and/or more accurate diagnosis, for the stratification of patients into prognostic classes, for disease monitoring, and/or for the development of specific and consequently more effective targeted therapies. Next-generation sequencing-based approaches, by providing both a high sequencing coverage and an unbiased view of complex systems without the need of an a priori knowledge of the targets of interest, have also imposed new analytical standards in these fields [2,7,10].

In the case of RNA studies for example, NGS-based approaches have overcome the use of microarrays: NGS allows the analysis of virtually all the RNA molecules, known and unknown, present in a sample, at a lower cost [2,7]. In addition, alternative splicing isoforms and long non-coding RNAs can also be highlighted [2,7,27–29], and specific small RNA classes can be enriched and sequenced [30–32]. Finally, recent applications are also showing the potential of NGS for single cells RNA sequencing [51,52]. Similarly, NGS has prompted the study of the epigenome and of the microbiome. By using the preparation protocols of specific libraries, it is possible to analyze the methylation status of DNA at a genome-wide level or by focusing on a custom set of genomic regions of interest [33,34]. Moreover, chromatin immunoprecipitation sequencing (ChIP-Seq) approaches have shown their efficacy in the study of the regulatory networks of gene expression at the genome-wide level, by allowing the identification of the targets of specific transcription factors [53]. Recently, the newest epigenomic methodologies, such as micrococcal nuclease sensitive sites sequencing (MNase-seq), DNase I hypersensitive sites sequencing (DNase-seq), formaldehyde-assisted isolation of regulatory elements sequencing (FAIRE-seq), and assay for transposase-accessible chromatin using sequencing (ATAC-seq), have been developed and have shown their reliability for the study of

chromatin accessibility at the genome-wide level to identify the epigenetic changes responsible for differential gene expression, cell proliferation, functional diversification and disease development [54].

Finally, by superseding the need of microbial cultivation, NGS-based techniques also gave a significant boost to metagenomics for the study of the microbial relationships with human physiology and pathology, and for the identification of specific microbial signatures related to a disease of interest [35–38,50]. It is now established that the human microbiome plays a role in healthy status acquisition and maintenance [50,55]. As a consequence, a microbial dysbiosis may contribute to diseases development and may provide novel actionable targets, not only for disease monitoring, but especially for the development of novel therapies [50].

As the costs of NGS continue to decrease, it is conceivable to hypothesize that these (and other) applications will become even more common and will be part of clinical practice. It has to be noticed that clinical and research studies require different pipelines; indeed, clinical studies need a lot of validation, remain rigid and are concerned only with findings that are actionable, whereas research studies are more fluid and concerned with discovery. Moreover, NGS technologies have accustomed us to rapid developments. For instance, an old limitation of NGS was its limited reads length, currently exceeded by an increased reads length capability on one side, and the availability of ad hoc designed assembly tools, on the other. Novel DNA sequencing techniques promise to further improve this aspect [2,7]. For example, nanopore-based sequencing chemistries have the double advantages of being able to avoid the amplification of libraries (and the related errors) and to allow the sequencing of very long reads (up to 950 kb) [56]. Once the accuracy of these methods is increased and the error-rate minimized, we will observe a novel revolution in the DNA sequencing field and a further reduction of the sequencing costs/genome [2].

Besides the above-mentioned improvements, in a similar manner, high-throughput mass spectrometry (MS) platforms have also been developed to exploit in depth the whole proteome and/or metabolome of cells, biological fluids and tissues [2,57]. Representing the final products of cellular processes, proteins and metabolites studies, also in combination with other -omic approaches, have the potential to further clarify pathogenetic mechanisms and highlight additional biomarkers. Mass spectrometry improvements allow the simultaneous analysis of multiple peptides/metabolites and also untargeted approaches for novel molecules detection [2,57]. However, peptides/metabolites identification is based on the comparison of the results obtained in the analyzed sample with respect to specific databases that still present limitations, especially for metabolomic evaluations. As further technological progress will be achieved, both proteomic and metabolomics profiling may integrate genomic data for a better diagnostic and prognostic classification. For more details regarding proteomic and metabolomics technologies, the interested reader is referred to specific papers on these topics [2,57].

### 3. Big Data Production, Big Data Analysis and Data Integration Methods

The term ‘big data’ indicates a huge amount of structured or unstructured data not analyzable by using traditional technologies, and characterized by great variety, high production speed, and extreme variability [58,59]. Technological advances in -omic sciences have brought them into the big data domain [60].

In recent years, we have been overwhelmed by a real technological escalation, far from the expected logarithmic trend based on Moore’s laws [2,39]. Currently, molecular protocols for high-throughput analyses have been well established, extensively validated and simplified to analyze an increasing number of samples in even less time and costs [2,7,10]. Liquid handling platforms for libraries preparation have also been developed to further optimize the samples preparation step by reducing both inter-samples variability and the hands-on time. The direct consequence of this phenomenon is that the more we are able to sequence fast, high-quantity bases at a high level of accuracy and lower cost, the more we accumulate data. The last few years have shown that our ability to generate data supersedes our possibility to analyze and interpret them [2,5–10]. Consider that a sequencing run is able to generate hundreds of gigabytes at a time and it has been estimated that in

the next ten years we may sequence up to two billion human genomes [61]. Considering the rapid development of novel technological platforms for data production, it is also possible that this number may be currently underestimated. A couple of years ago, an interesting report already compared genomics to other big data classical domains, and considering four typical features of a dataset's life-cycle (i.e., acquisition, storage, distribution, and analysis), defined genomic as a "four-headed beast" since its needs overcome that of all the others [60].

This big data production has imposed the development and validation of specific bioinformatic tools for their analysis, starting from quality check, background noise minimization and reads normalization [62]. Data analysis and interpretation now represent the most important challenge to be addressed when approaching high-throughput analyses. Today, NGS methods offer a plethora of applications investigating different biological systems, both in a deeper focused and/or genome-wide manner [2,5–10,62]. Each of these procedures requires specific validated pipelines to address a specific biological question, i.e., identify disease-related mutations, obtain a differential expression analysis, or define the microbiome composition [2,5–10,62,63]. These operations not only need highly qualified, specific expertise to develop highly sensitive (and preferably easy-to-use) tools for data management and analysis, but also need cooperative efforts to establish quality guidelines ensuring data comparison among different datasets and different laboratories worldwide [9,64]. Indeed, if each laboratory uses its own pipeline to analyze its own results, the risk of finding everything and its exact opposite, without the possibility of comparing the results obtained by different studies, may become a reality. Reproducibility is a key feature of scientific research; high-throughput data are challenging in this regard due to the high variability of the samples analyzed and/or of the experimental procedures, and the complexity of the data and the use of not properly validated and/or standardized pipelines. Statistician-derived methods may be useful in this context by supporting experimental design and reproducibility, preprocessing, structure learning, and data integration [65]. The information is in the data: the methodology for their correct interpretation must be widely validated and standardized to ensure laboratory data harmonization and be sure that significant differences in a specific sample, or in a population, are really due to a relevant biological alteration and not to biases attributable to the used analytical approaches (both at molecular and bioinformatics levels). The re-analysis with updated pipelines of samples previously reported as "negative", and the management of the so-called incidental findings are other relevant hot topics in this field [42,66,67]. Both of these aspects also require caution and shared guidelines.

In addition, the continued development of novel NGS-based strategies requires the continued availability of ad hoc pipelines able to overcome possible methodological limitations and highlight the biologically relevant information. For example, the recent introduction of methodologies able to achieve a single cell resolution gave an unprecedented opportunity to study cellular heterogeneity at different levels. This, in turn, requires computational methods able to overcome some criticisms, such as the systematic noise, the complexity of the data, and the need of validation techniques [68,69]. Despite several computational solutions, an important bottleneck is currently represented by the need to overcome biological and technical variability in single cell data [69]. Since a number of papers addressing these issues are being published, it is conceivable to suppose that current challenges and limitations in single cell analysis will soon be overcome, opening the way to new applications and opportunities [70–72].

Furthermore, specific comprehensive databases are required to compare the results obtained in the samples of interest and infer biologically and clinically relevant information [2,5–10,62,64]. The continued evolution of NGS applications also requires the corresponding and continued evolution of bioinformatic instruments (including the update of the reference databases). Limitations in the reference databases also negatively impact the interpretations of data derived from genome-wide proteomic or metabolomics analyses, whose potential seems now to be, consequently, underestimated [2,57]. In addition, the increasing throughput of the sequencers imposes the need for tools that are able to manage a huge amount of data. Consequently, an additional problem

is the storage of the generated data. Despite the technological advances in big data production, their management (including data storage and the computational resources required for their analysis and interpretation) is still expensive [73]. A promising solution for -omics data handling is represented by cloud computing. These systems are based on virtual, web-based solutions, able to use multiple computational resources simultaneously, escalating the computational power in respect to local-based servers [73]. The large diffusion of cloud computing is partly due to the availability of infrastructures, such as Hadoop [74]. Hadoop is an open-source software that allows the processing of large datasets by distributing them across multiple computer nodes; thus, it is particularly suitable for bioinformatic purposes [75]. Depending on the services offered, in terms of the level of functionality given to the user by the cloud provider, cloud computing-based solutions for big data manipulation can be classified as Data as a Service (DaaS), Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) [75]. In addition to integrating data and specific tools for their analysis, the bioinformatics cloud should provide technologies for high-speed data transfer, allow the development of customized pipelines by the users, and be publicly accessible. Further, once data have been properly stored, we need a system to manage them in order not only to efficiently archive data, but also to make them easily available to the user on demand. In particular, different kinds of NoSQL (Not only Structured Query Language) databases have rapidly emerged showing their benefits over traditional relational databases in the speed of data storage, indexing, and query retrieval [76]. Finally, LIMS (laboratory information management system) solutions support standardized data management and tracking systems [77–79]. Large repositories, such as TCGA [80] and CBioPortal [81], represent successful examples regarding big data analysis, management, integration, and sharing.

While a lot of companies are offering cloud-based solutions for data storage, ethical concerns are emerging regarding their safety and properties [82]. Similar issues are also emerging with regard to data sharing [2,10]. The sharing of research data offers the unique opportunity to increase knowledge by avoiding unnecessary duplications and obtaining novel, useful information from the re-analysis of the same datasets. However, it imposes several challenges of ethical, cultural, legal, financial, and technical nature [83]. Even if a couple of years ago the Regulatory and Ethics Working Group of the Global Alliance for Genomics and Health proposed a standardized model for data sharing [84], its application seems far away.

Considering the escalation of high-throughput technologies and their exponentially increasing ability to generate high-quality datasets of huge size, big data algorithms and high-performance computing (HPC) systems are needed for large-scale analyses [59,61,73,85]. Schmidt et al. recently reviewed big data analysis techniques for NGS data and HPC solutions and hypothesized a shift from model-driven to data-driven science [61]. This assumption requires caution since, if brought to its extreme consequences, the risk is to look just at data and not at biological systems.

Another important issue is related to the need of instruments for data integration. Indeed, big data production related to -omic sciences has led to the onset of the so-called 'systems biology'. Systems biology proposes a holistic view based on the integration of multidisciplinary data to infer mechanistic associations and gain insights into biologic processes, including complex human diseases, in view of personalized medicine [73]. However, the high-throughput technologies described above are often focused on a specific -omic network (i.e., DNA, RNA, proteins, or metabolites). Consequently, even if these approaches are very effective, they miss a comprehensive view since they focus on a single system at a time. This limitation may contribute to the gap between the ability to produce huge amounts of data and the difficulty to correlate these data to complex phenotypes and to predict their outcomes. Since high-throughput analyses coupled with bioinformatics are becoming routine procedures, it is easy to hypothesize that the future direction will be the integration of multi-omics data. By integrating multiple sources of data, it is possible to highlight information that may be underestimated in the analysis of just one -omic level, and it is possible to reinforce the strength of some associations by confirming them at multiple levels [86]. The identification of specific interactions across -omic levels may shed light on the molecular mechanisms underlying human diseases and support the



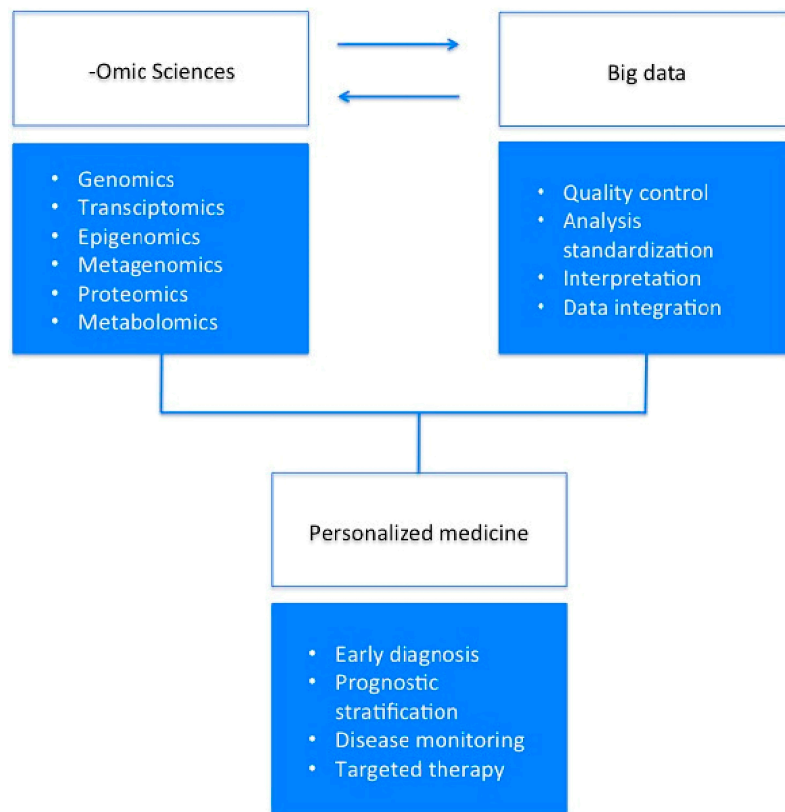
identification of biomarkers for disease risk prediction or monitoring. However, this process is currently still complex and difficult due to several factors, i.e., different sources of data and different data formats, the lack or redundancy of databases, and the lack of data standards [87]. Thus, different strategies are being proposed to efficiently integrate the information concerning the complex relationships across multi-omics levels by using systems genomic approaches [86,88–93]. Based on the algorithms used, data integration methods can be classified into three categories: unsupervised, supervised, and semi-supervised [93]. Briefly, unsupervised methods use different approaches to cluster objects into different categories based on their biological profiles; supervised methods start from known labels (i.e., disease or healthy) to predict the related patterns and assign the unlabeled data to each of them; and, finally, semi-supervised methods, often graph-based, use both labeled and unlabeled samples to develop learning algorithms based on similarity networks [93]. In this context, Yugi et al., proposed a “trans-omic” analysis to obtain a global network from multi-omics data and applied it to three case-studies showing the potentialities of this approach, even if technological improvements, including validation strategies, are required [90]. Dimension reduction approaches have also been used for the analysis of multiple data sets [92]. These methods calculate the most valuable linear relationships able to explain the correlation across data sets and can also evaluate the variability effects of outliers [92].

Machine learning and system genomics (MLSG)-based approaches are also showing their reliability in the identification of genotype–phenotype relationships, as a result of the integration of multiple data from multi-omics analyses by using predictive algorithms and data mining [94,95]. Indeed, machine learning employs predictive algorithms able to recognize specific patterns from complex data and to learn from them in order to use this knowledge to make reliable predictions. Thus, this kind of analysis requires as its first step the design of a model, starting from known datasets used as examples to generate the associations, before it can be used to make predictions [94,95]. In the field of MLSG, different software for the prediction of genotype–phenotype relationships from multi-omics data have been developed based on different methods, i.e., different models to predict significant associations, showing the potential of MLSG in predicting diseases outcome [94]. Despite this great potential as a key technology in supporting clinical decisions, machine learning is still not widely used [96]. This is partly due to the lack of the required skills and expertise of life sciences researchers that are often unable to infer biologically relevant information from the huge amount of data they produce. The need for multidisciplinary teams has already been postulated in recent years. Alyass et al., for example, pointed out that the road to personalized medicine requires a strong, or we can say a revolutionary, integration between traditionally well-separated disciplines [73]. This is also the reason why many efforts are being made to make computational instruments easy-to-use for inexperienced users. Luo et al. developed Automated Machine-Learning (Auto-ML) to automate the whole machine learning process and support patients’ outcome predictions [96]. In this context, graphical user interfaces (GUI) may support inexperienced users in respect to a command line interface (CLI) [97].

Based on all the above, it is easy to predict that the next few years will feature further development of statistical, mathematical and information technology (IT) instruments in the -omic context. This will completely change the care process and the concept of medicine, and will also require careful regulation to avoid the risks related to a data-centric view.

#### 4. Conclusions

Since it has been estimated that the cost of the sequencing will continue to decrease, whole-genome sequencing may become a routine clinical practice to obtain clinically relevant information for the correct and early diagnosis, to determine the most proper therapy, and for disease monitoring. The integration of these data with those derived from other -omic approaches will shed light on the mechanisms underlying human diseases and will allow the identification of novel biomarkers for the diagnosis and monitoring of diseases, as well as actionable targets for specific therapies in view of an even more personalized medicine approach (Figure 1).



**Figure 1.** The integration between different -omic sciences and validated and standardized tools for big data analysis will bring personalized medicine into real clinical practice.

Personalized medicine means that medical decisions are customized to each individual based on specific biomarkers obtained from multi-omics data.

The availability of high-throughput methods coupled with tools for big data analysis and integration, and machine learning-based approaches has the potential to bring personalized medicine into real medical practice. Indeed, the access to an individual's genomic content will provide information not only on the underlying disease but also may highlight actionable target for specific therapies, and infer prediction regarding specific outcomes. To bring this model into clinical practice, issues concerning big data production and interpretation need to be assessed. Once we are able to validate and establish shared pipelines for the accurate analysis of high-throughput-derived data, also including the ethical aspects to regulate privacy issues and data sharing, we may be able to fight the data struggle.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Yadav, S.P. The wholeness in suffix -omics, -omes, and the word om. *J. Biomol. Tech.* **2007**, *18*, 277. [[PubMed](#)]
2. Sandhu, C.; Qureshi, A.; Emili, A. Panomics for Precision Medicine. *Trends Mol. Med.* **2017**. [[CrossRef](#)] [[PubMed](#)]
3. Venter, J.C.; Adams, M.D.; Myers, E.W.; Li, P.W.; Mural, R.J.; Sutton, G.G.; Smith, H.O.; Yandell, M.; Evans, C.A.; Holt, R.A.; et al. The sequence of the human genome. *Science* **2001**, *291*, 1304–1351. [[CrossRef](#)] [[PubMed](#)]
4. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **2004**, *431*, 931–945.

5. Mardis, E.R. The impact of next-generation sequencing technology on genetics. *Trends Genet.* **2008**, *24*, 133–141. [[CrossRef](#)] [[PubMed](#)]
6. Reuter, J.A.; Spacek, D.V.; Snyder, M.P. High-Throughput Sequencing Technologies. *Mol. Cell* **2015**, *58*, 586–597. [[CrossRef](#)] [[PubMed](#)]
7. Precone, V.; Del Monaco, V.; Esposito, M.V.; De Palma, F.D.; Ruocco, A.; Salvatore, F.; D’Argenio, V. Cracking the Code of Human Diseases Using Next-Generation Sequencing: Applications, Challenges, and Perspectives. *Biomed. Res. Int.* **2015**, 161648. [[CrossRef](#)] [[PubMed](#)]
8. Kulkarni, P.; Frommolt, P. Challenges in the Setup of Large-scale Next-Generation Sequencing Analysis Workflows. *Comput. Struct. Biotechnol. J.* **2017**, *15*, 471–477. [[CrossRef](#)] [[PubMed](#)]
9. Roy, S.; Coldren, C.; Karunamurthy, A.; Kip, N.S.; Klee, E.W.; Lincoln, S.E.; Leon, A.; Pullambhatla, M.; Temple-Smolkin, R.L.; Voelkerding, K.V.; et al. Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. *J. Mol. Diagn.* **2018**, *20*, 4–27. [[CrossRef](#)] [[PubMed](#)]
10. Caspar, S.M.; Dubacher, N.; Kopps, A.M.; Meienberg, J.; Henggeler, C.; Matyas, G. Clinical sequencing: From raw data to diagnosis with lifetime value. *Clin. Genet.* **2018**, *93*, 508–519. [[CrossRef](#)] [[PubMed](#)]
11. Sanger, F.; Nicklen, S.; Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463–5467. [[CrossRef](#)] [[PubMed](#)]
12. D’Argenio, V.; Notomista, E.; Petrillo, M.; Cantiello, P.; Cafaro, V.; Izzo, V.; Naso, B.; Cozzuto, L.; Durante, L.; Troncone, L.; et al. Complete sequencing of *Novosphingobium* sp. PP1Y reveals a biotechnologically meaningful metabolic pattern. *BMC Genom.* **2014**, *15*, 384. [[CrossRef](#)] [[PubMed](#)]
13. D’Argenio, V.; Petrillo, M.; Pasanisi, D.; Pagliarulo, C.; Colicchio, R.; Talà, A.; de Biase, M.S.; Zanfardino, M.; Scolamiero, E.; Pagliuca, C.; et al. The complete 12 Mb genome and transcriptome of *Nonomuraea gerenzanensis* with new insights into its duplicated “magic” RNA polymerase. *Sci. Rep.* **2016**, *6*, 18. [[CrossRef](#)] [[PubMed](#)]
14. Horai, M.; Mishima, H.; Hayashida, C.; Kinoshita, A.; Nakane, Y.; Matsuo, T.; Tsuruda, K.; Yanagihara, K.; Sato, S.; Imanishi, D.; et al. Detection of de novo single nucleotide variants in offspring of atomic-bomb survivors close to the hypocenter by whole-genome sequencing. *J. Hum. Genet.* **2017**. [[CrossRef](#)] [[PubMed](#)]
15. Jun, G.; Manning, A.; Almeida, M.; Zawistowski, M.; Wood, A.R.; Teslovich, T.M.; Fuchsberger, C.; Feng, S.; Cingolani, P.; Gaulton, K.J.; et al. Evaluating the contribution of rare variants to type 2 diabetes and related traits using pedigrees. *Proc. Natl. Acad. Sci. USA* **2017**. [[CrossRef](#)] [[PubMed](#)]
16. Weisz Hubshman, M.; Broekman, S.; van Wijk, E.; Cremers, F.; Abu-Diab, A.; Samer, K.; Tzur, S.; Lagovsky, I.; Smirin-Yosef, P.; Sharon, D.; et al. Whole-exome sequencing reveals POC5 as a novel gene associated with autosomal recessive retinitis pigmentosa. *Hum. Mol. Genet.* **2017**. [[CrossRef](#)] [[PubMed](#)]
17. Tada, H.; Inaba, S.; Pozharitckaia, D.; Kawashiri, M.A. Prominent Tendon Xanthomas and Abdominal Aortic Aneurysm Associated with Cerebrotendinous Xanthomatosis Identified Using Whole Exome Sequencing. *Intern. Med.* **2017**. [[CrossRef](#)] [[PubMed](#)]
18. Calhoun, J.D.; Vanoye, C.G.; Kok, F.; George, A.L., Jr.; Kearney, J.A. Characterization of a *KCNB1* variant associated with autism, intellectual disability, and epilepsy. *Neurol. Genet.* **2017**, *3*, e198. [[CrossRef](#)] [[PubMed](#)]
19. D’Argenio, V.; Frisso, G.; Precone, V.; Boccia, A.; Fienga, A.; Pacileo, G.; Limongelli, G.; Paoletta, G.; Calabrò, R.; Salvatore, F. DNA sequence capture and next-generation sequencing for the molecular diagnosis of genetic cardiomyopathies. *J. Mol. Diagn.* **2014**, *16*, 32–44. [[CrossRef](#)] [[PubMed](#)]
20. Miller, E.M.; Patterson, N.E.; Zechmeister, J.M.; Bejerano-Sagie, M.; Delio, M.; Patel, K.; Ravi, N.; Quispe-Tintaya, W.; Maslov, A.; Simmons, N.; et al. Development and validation of a targeted next generation DNA sequencing panel outperforming whole exome sequencing for the identification of clinically relevant genetic variants. *Oncotarget* **2017**, *8*, 102033–102045. [[CrossRef](#)] [[PubMed](#)]
21. Kalsner, L.; Twachtman-Bassett, J.; Tokarski, K.; Stanley, C.; Dumont-Mathieu, T.; Cotney, J.; Chamberlain, S. Genetic testing including targeted gene panel in a diverse clinical population of children with autism spectrum disorder: Findings and implications. *Mol. Genet. Genom. Med.* **2017**. [[CrossRef](#)] [[PubMed](#)]
22. D’Argenio, V.; Esposito, M.V.; Telese, A.; Precone, V.; Starnone, F.; Nunziato, M.; Cantiello, P.; Iorio, M.; Evangelista, E.; D’Aiuto, M.; et al. The molecular analysis of BRCA1 and BRCA2: Next-generation sequencing supersedes conventional approaches. *Clin. Chim. Acta* **2015**, *446*, 221–225. [[CrossRef](#)] [[PubMed](#)]



23. Trujillano, D.; Weiss, M.E.; Köster, J.; Papachristos, E.B.; Werber, M.; Kandaswamy, K.K.; Marais, A.; Eichler, S.; Creed, J.; Baysal, E.; et al. Validation of a semiconductor next-generation sequencing assay for the clinical genetic screening of *CFTR*. *Mol. Genet. Genom. Med.* **2015**, *3*, 396–403. [[CrossRef](#)] [[PubMed](#)]
24. Esposito, M.V.; Nunziato, M.; Starnone, F.; Telese, A.; Calabrese, A.; D’Aiuto, G.; Pucci, P.; D’Aiuto, M.; Baralle, F.; D’Argenio, V.; et al. A Novel Pathogenic BRCA1 Splicing Variant Produces Partial Intron Retention in the Mature Messenger RNA. *Int. J. Mol. Sci.* **2016**, *17*, 2145. [[CrossRef](#)] [[PubMed](#)]
25. Nunziato, M.; Starnone, F.; Lombardo, B.; Pensabene, M.; Condello, C.; Verdesca, F.; Carlomagno, C.; De Placido, S.; Pastore, L.; Salvatore, F.; et al. Fast Detection of a *BRCA2* Large Genomic Duplication by Next Generation Sequencing as a Single Procedure: A Case Report. *Int. J. Mol. Sci.* **2017**, *18*, 2487. [[CrossRef](#)] [[PubMed](#)]
26. Xu, Y.; Wang, H.; Xiao, B.; Wei, W.; Liu, Y.; Ye, H.; Ying, X.M.; Chen, Y.W.; Liu, X.Q.; Ji, X.; et al. Novel noncontiguous duplications identified with a comprehensive mutation analysis in the *DMD* gene by *DMD* gene-targeted sequencing. *Gene* **2017**. [[CrossRef](#)] [[PubMed](#)]
27. Panagopoulos, I.; Gorunova, L.; Spetalen, S.; Bassarova, A.; Beiske, K.; Micci, F.; Heim, S. Fusion of the genes ataxin 2 like, *ATXN2L*, and Janus kinase 2, *JAK2*, in cutaneous CD4 positive T-cell lymphoma. *Oncotarget* **2017**, *8*, 103775–103784. [[CrossRef](#)] [[PubMed](#)]
28. Su, Y.T.; Chen, R.; Wang, H.; Song, H.; Zhang, Q.; Chen, L.Y.; Lappin, H.; Vasconcelos, G.; Lita, A.; Maric, D.; et al. Novel Targeting of Transcription and Metabolism in Glioblastoma. *Clin. Cancer Res.* **2017**. [[CrossRef](#)] [[PubMed](#)]
29. Chen, B.; Jiang, L.; Zhong, M.L.; Li, J.F.; Li, B.S.; Peng, L.J.; Dai, Y.T.; Cui, B.W.; Yan, T.Q.; Zhang, W.N.; et al. Identification of fusion genes and characterization of transcriptome features in T-cell acute lymphoblastic leukemia. *Proc. Natl. Acad. Sci. USA* **2017**. [[CrossRef](#)] [[PubMed](#)]
30. Aceto, S.; Sica, M.; De Paolo, S.; D’Argenio, V.; Cantiello, P.; Salvatore, F.; Gaudio, L. The analysis of the inflorescence miRNome of the orchid *Orchis italica* reveals a *DEF*-like MADS-box gene as a new miRNA target. *PLoS ONE* **2014**, *9*, e97839. [[CrossRef](#)] [[PubMed](#)]
31. Nardelli, C.; Granata, I.; Iaffaldano, L.; D’Argenio, V.; Del Monaco, V.; Maruotti, G.M.; Omodei, D.; Del Vecchio, L.; Martinelli, P.; Salvatore, F.; et al. miR-138/miR-222 Overexpression Characterizes the miRNome of Amniotic Mesenchymal Stem Cells in Obesity. *Stem Cells Dev.* **2017**, *26*, 4–14. [[CrossRef](#)] [[PubMed](#)]
32. D’Argenio, V.; Del Monaco, V.; Paparo, L.; De Palma, F.D.E.; Nocerino, R.; D’Alessio, F.; Visconte, F.; Discepolo, V.; Del Vecchio, L.; Salvatore, F.; et al. Altered miR-193a-5p expression in children with cow’s milk allergy. *Allergy* **2017**. [[CrossRef](#)]
33. Pu, W.; Wang, C.; Chen, S.; Zhao, D.; Zhou, Y.; Ma, Y.; Wang, Y.; Li, C.; Huang, Z.; Jin, L.; et al. Targeted bisulfite sequencing identified a panel of DNA methylation-based biomarkers for esophageal squamous cell carcinoma (ESCC). *Clin. Epigenet.* **2017**, *9*, 129. [[CrossRef](#)] [[PubMed](#)]
34. Widschwendter, M.; Evans, I.; Jones, A.; Ghazali, S.; Reisel, D.; Ryan, A.; Gentry-Maharaj, A.; Zikan, M.; Cibula, D.; Eichner, J.; et al. Methylation patterns in serum DNA for early identification of disseminated breast cancer. *Genome Med.* **2017**, *9*, 115. [[CrossRef](#)] [[PubMed](#)]
35. D’Argenio, V.; Precone, V.; Casaburi, G.; Miele, E.; Martinelli, M.; Staiano, A.; Salvatore, F.; Sacchetti, L. An altered gut microbiome profile in a child affected by Crohn’s disease normalized after nutritional therapy. *Am. J. Gastroenterol.* **2013**, *108*, 851–852. [[CrossRef](#)] [[PubMed](#)]
36. D’Argenio, V.; Casaburi, G.; Precone, V.; Pagliuca, C.; Colicchio, R.; Sarnataro, D.; Discepolo, V.; Kim, S.M.; Russo, I.; Del Vecchio Blanco, G.; et al. Metagenomics Reveals Dysbiosis and a Potentially Pathogenic *N. flavescens* Strain in Duodenum of Adult Celiac Patients. *Am. J. Gastroenterol.* **2016**, *111*, 879–890. [[CrossRef](#)] [[PubMed](#)]
37. D’Argenio, V.; Casaburi, G.; Precone, V.; Pagliuca, C.; Colicchio, R.; Sarnataro, D.; Discepolo, V.; Kim, S.M.; Russo, I.; Del Vecchio Blanco, G.; et al. No Change in the Mucosal Gut Microbiome is Associated with Celiac Disease-Specific Microbiome Alteration in Adult Patients. *Am. J. Gastroenterol.* **2016**, *111*, 1659–1661. [[CrossRef](#)] [[PubMed](#)]
38. D’Argenio, V.; Torino, M.; Precone, V.; Casaburi, G.; Esposito, M.V.; Iaffaldano, L.; Malapelle, U.; Troncone, G.; Coto, I.; Cavalcanti, P.; et al. The Cause of Death of a Child in the 18th Century Solved by Bone Microbiome Typing Using Laser Microdissection and Next Generation Sequencing. *Int. J. Mol. Sci.* **2017**, *18*, 109. [[CrossRef](#)] [[PubMed](#)]

39. Hayden, E.C. Technology: The \$1000 genome. *Nature* **2014**, *507*, 294–295. [[CrossRef](#)] [[PubMed](#)]
40. Sanna, V.; Zarrilli, F.; Nardiello, P.; D'Argenio, V.; Rocino, A.; Coppola, A.; DI Minno, G.; Castaldo, G. Mutational spectrum of F8 gene and prothrombotic gene variants in haemophilia A patients from Southern Italy. *Haemophilia* **2008**, *14*, 796–803. [[CrossRef](#)] [[PubMed](#)]
41. Larsen, M.; Rost, S.; El Hajj, N.; Ferbert, A.; Deschauer, M.; Walter, M.C.; Schoser, B.; Tacik, P.; Kress, W.; Müller, C.R. Diagnostic approach for FSHD revisited: *SMCHD1* mutations cause FSHD2 and act as modifiers of disease severity in FSHD1. *Eur. J. Hum. Genet.* **2015**, *23*, 808–816. [[CrossRef](#)] [[PubMed](#)]
42. Weber, S.; Büscher, A.K.; Hagmann, H.; Liebau, M.C.; Heberle, C.; Ludwig, M.; Rath, S.; Alberer, M.; Beissert, A.; Zenker, M.; et al. Dealing with the incidental finding of secondary variants by the example of SRNS patients undergoing targeted next-generation sequencing. *Pediatr. Nephrol.* **2016**, *31*, 73–81. [[CrossRef](#)] [[PubMed](#)]
43. Maxwell, S.M.; Colls, P.; Hodes-Wertz, B.; McCulloh, D.H.; McCaffrey, C.; Wells, D.; Munné, S.; Grifo, J.A. Why do euploid embryos miscarry? A case-control study comparing the rate of aneuploidy within presumed euploid embryos that resulted in miscarriage or live birth using next-generation sequencing. *Fertil. Steril.* **2016**, *106*, 1414–1419. [[CrossRef](#)] [[PubMed](#)]
44. D'Argenio, V.; Nunziato, M.; D'Uonno, N.; Borrillo, F.; Vallone, R.; Conforti, A.; De Rosa, P.; Tomaiuolo, R.; Cariati, F. Indications and limitations for preimplantation genetic diagnosis. *Biochim. Clin.* **2017**, *41*, 314–321.
45. Huang, C.E.; Ma, G.C.; Jou, H.J.; Lin, W.H.; Lee, D.J.; Lin, Y.S.; Ginsberg, N.A.; Chen, H.F.; Chang, F.M.; Chen, M. Noninvasive prenatal diagnosis of fetal aneuploidy by circulating fetal nucleated red blood cells and extravillous trophoblasts using silicon-based nanostructured microfluidics. *Mol. Cytogenet.* **2017**, *10*, 44. [[CrossRef](#)] [[PubMed](#)]
46. Harper, J.C.; Aittomäki, K.; Borry, P.; Cornel, M.C.; de Wert, G.; Dondorp, W.; Geraedts, J.; Gianaroli, L.; Ketterson, K.; Liebaers, I.; et al. Recent developments in genetics and medically assisted reproduction: From research to clinical applications. *Eur. J. Hum. Genet.* **2017**. [[CrossRef](#)]
47. D'Argenio, V.; Tomaiuolo, R.; Cariati, F. Whole genome amplification on single cell. *Biochim. Clin.* **2016**, *40*, 293–301.
48. Liu, H.E.; Triboulet, M.; Zia, A.; Vuppapalaty, M.; Kidess-Sigal, E.; Collier, J.; Natu, V.S.; Shokoohi, V.; Che, J.; Renier, C.; et al. Workflow optimization of whole genome amplification and targeted panel sequencing for CTC mutation detection. *NPJ Genom. Med.* **2017**, *2*. [[CrossRef](#)] [[PubMed](#)]
49. Müller, S.; Kohanbash, G.; Liu, S.J.; Alvarado, B.; Carrera, D.; Bhaduri, A.; Watchmaker, P.B.; Yagnik, G.; Di Lullo, E.; Malatesta, M.; et al. Single-cell profiling of human gliomas reveals macrophage ontogeny as a basis for regional differences in macrophage activation in the tumor microenvironment. *Genome Biol.* **2017**, *18*, 234. [[CrossRef](#)] [[PubMed](#)]
50. D'Argenio, V.; Salvatore, F. The role of the gut microbiome in the healthy adult status. *Clin. Chim. Acta* **2015**, *451*, 97–102. [[CrossRef](#)] [[PubMed](#)]
51. Yuzwa, S.A.; Borrett, M.J.; Innes, B.T.; Voronova, A.; Ketela, T.; Kaplan, D.R.; Bader, G.D.; Miller, F.D. Developmental Emergence of Adult Neural Stem Cells as Revealed by Single-Cell Transcriptional Profiling. *Cell Rep.* **2017**, *21*, 3970–3986. [[CrossRef](#)] [[PubMed](#)]
52. Zong, C.C. Single-cell RNA-seq study determines the ontogeny of macrophages in glioblastomas. *Genome Biol.* **2017**, *18*, 235. [[CrossRef](#)] [[PubMed](#)]
53. Pavesi, G. ChIP-Seq Data Analysis to Define Transcriptional Regulatory Networks. *Adv. Biochem. Eng. Biotechnol.* **2017**, *160*, 1–14. [[PubMed](#)]
54. Buenostro, J.D.; Wu, B.; Chang, H.Y.; Greenleaf, W.J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* **2015**, *109*, 1–9.
55. Perez-Muñoz, M.E.; Arrieta, M.C.; Ramer-Tait, A.E.; Walter, J. A critical assessment of the “sterile womb” and “in utero colonization” hypotheses: Implications for research on the pioneer infant microbiome. *Microbiome* **2017**, *5*, 48. [[CrossRef](#)] [[PubMed](#)]
56. Tyson, J.R.; O'Neil, N.J.; Jain, M.; Olsen, H.E.; Hieter, P.; Snutch, T.P. MinION-based long-read sequencing and assembly extends the *Caenorhabditis elegans* reference genome. *Genome Res.* **2017**. [[CrossRef](#)]
57. Jacob, M.; Lopata, A.L.; Dasouki, M.; Abdel Rahman, A.M. Metabolomics toward personalized medicine. *Mass Spectrom. Rev.* **2017**. [[CrossRef](#)] [[PubMed](#)]
58. Baro, E.; Degoul, S.; Beuscart, R.; Chazard, E. Toward a Literature-Driven Definition of Big Data in Healthcare. *Biomed. Res. Int.* **2015**, *2015*, 639021. [[CrossRef](#)] [[PubMed](#)]

59. Yang, A.; Troup, M.; Ho, J.W.K. Scalability and Validation of Big Data Bioinformatics Software. *Comput. Struct. Biotechnol. J.* **2017**, *15*, 379–386. [[CrossRef](#)] [[PubMed](#)]
60. Stephens, Z.D.; Lee, S.Y.; Faghri, F.; Campbell, R.H.; Zhai, C.; Efron, M.J.; Iyer, R.; Schatz, M.C.; Sinha, S.; Robinson, G.E. Big Data: Astronomical or Genomical? *PLoS Biol.* **2015**, *13*, e1002195. [[CrossRef](#)] [[PubMed](#)]
61. Schmidt, B.; Hildebrandt, A. Next-generation sequencing: Big data meets high performance computing. *Drug Discov. Today* **2017**, *22*, 712–717. [[CrossRef](#)] [[PubMed](#)]
62. Yin, Z.; Lan, H.; Tan, G.; Lu, M.; Vasilakos, A.V.; Liu, W. Computing Platforms for Big Biological Data Analytics: Perspectives and Challenges. *Comput. Struct. Biotechnol. J.* **2017**, *15*, 403–411. [[CrossRef](#)] [[PubMed](#)]
63. D'Argenio, V. Human Microbiome Acquisition and Bioinformatic Challenges in Metagenomic Studies. *Int. J. Mol. Sci.* **2018**, *19*, 383. [[CrossRef](#)] [[PubMed](#)]
64. Mason, C.E.; Afshinnekoo, E.; Tighe, S.; Wu, S.; Levy, S. International Standards for Genomes, Transcriptomes, and Metagenomes. *J. Biomol. Tech.* **2017**, *28*, 8–18. [[CrossRef](#)] [[PubMed](#)]
65. Morris, J.S.; Baladandayuthapani, V. Statistical Contributions to Bioinformatics: Design, Modeling, Structure Learning, and Integration. *Stat. Model.* **2017**, *17*, 245–289. [[CrossRef](#)] [[PubMed](#)]
66. Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W.W.; Hegde, M.; Lyon, E.; Spector, E.; et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **2015**, *17*, 405–424. [[CrossRef](#)] [[PubMed](#)]
67. Hoskinson, D.C.; Dubuc, A.M.; Mason-Suares, H. The current state of clinical interpretation of sequence variants. *Curr. Opin. Genet. Dev.* **2017**, *42*, 33–39. [[CrossRef](#)] [[PubMed](#)]
68. Yuan, G.C.; Cai, L.; Elowitz, M.; Enver, T.; Fan, G.; Guo, G.; Irizarry, R.; Kharchenko, P.; Kim, J.; Orkin, S.; et al. Challenges and emerging directions in single-cell analysis. *Genome Biol.* **2017**, *18*, 84. [[CrossRef](#)] [[PubMed](#)]
69. Fiers, M.W.E.J.; Minnoye, L.; Aibar, S.; Bravo González-Blas, C.; Kalender Atak, Z.; Aerts, S. Mapping gene regulatory networks from single-cell omics data. *Brief. Funct. Genom.* **2018**. [[CrossRef](#)] [[PubMed](#)]
70. Risso, D.; Perraudeau, F.; Gribkova, S.; Dudoit, S.; Vert, J.P. A general and flexible method for signal extraction from single-cell RNA-seq data. *Nat. Commun.* **2018**, *9*, 284. [[CrossRef](#)] [[PubMed](#)]
71. Sinha, D.; Kumar, A.; Kumar, H.; Bandyopadhyay, S.; Sengupta, D. dropClust: Efficient clustering of ultra-large scRNA-seq data. *Nucleic Acids Res.* **2018**. [[CrossRef](#)] [[PubMed](#)]
72. Wolf, F.A.; Angerer, P.; Theis, F.J. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **2018**, *19*, 15. [[CrossRef](#)] [[PubMed](#)]
73. Alyass, A.; Turcotte, M.; Meyre, D. From big data analysis to personalized medicine for all: Challenges and opportunities. *BMC Med. Genom.* **2015**, *8*, 33. [[CrossRef](#)] [[PubMed](#)]
74. Apache Hadoop. Available online: <http://hadoop.apache.org> (accessed on 16 February 2018).
75. Dai, L.; Gao, X.; Guo, Y.; Xiao, J.; Zhang, Z. Bioinformatics clouds for big data manipulation. *Biol. Direct* **2012**, *7*, 43. [[CrossRef](#)] [[PubMed](#)]
76. Schulz, W.L.; Nelson, B.G.; Felker, D.K.; Durant, T.J.S.; Torres, R. Evaluation of relational and NoSQL database architectures to manage genomic annotations. *J. Biomed. Inform.* **2016**, *64*, 288–295. [[CrossRef](#)] [[PubMed](#)]
77. Calabria, A.; Spinozzi, G.; Benedicenti, F.; Tenderini, E.; Montini, E. adLIMS: A customized open source software that allows bridging clinical and basic molecular research studies. *BMC Bioinform.* **2015**, *16*, S5. [[CrossRef](#)] [[PubMed](#)]
78. Chen, Y.; Lin, Y.; Yuan, X.; Shen, B. LIMS and Clinical Data Management. *Adv. Exp. Med. Biol.* **2016**, *939*, 225–239. [[PubMed](#)]
79. Craig, T.; Holland, R.; D'Amore, R.; Johnson, J.R.; McCue, H.V.; West, A.; Zulkower, V.; Tekotte, H.; Cai, Y.; Swan, D.; et al. Leaf LIMS: A Flexible Laboratory Information Management System with a Synthetic Biology Focus. *ACS Synth. Biol.* **2017**, *6*, 2273–2280. [[CrossRef](#)] [[PubMed](#)]
80. The Cancer Genome Atlas. Available online: <https://tcga-data.nci.nih.gov/tcga/> (accessed on 16 February 2018).
81. cBIOPortal for Cancer Genomics. Available online: <http://www.cbiportal.org> (accessed on 16 February 2018).
82. Tang, H.; Jiang, X.; Wang, X.; Wang, S.; Sofia, H.; Fox, D.; Lauter, K.; Malin, B.; Telenti, A.; Xiong, L.; et al. Protecting genomic data analytics in the cloud: state of the art and opportunities. *BMC Med. Genom.* **2016**, *9*, 63. [[CrossRef](#)] [[PubMed](#)]
83. Figueiredo, A.S. Data Sharing: Convert Challenges into Opportunities. *Front. Public Health* **2017**, *5*, 327. [[CrossRef](#)] [[PubMed](#)]

84. Kosseim, P.; Dove, E.S.; Baggaley, C.; Meslin, E.M.; Cate, F.H.; Kaye, J.; Harris, J.R.; Knoppers, B.M. Building a data sharing model for global genomic research. *Genome Biol.* **2014**, *15*. [[CrossRef](#)] [[PubMed](#)]
85. Raja, K.; Patrick, M.; Gao, Y.; Madu, D.; Yang, Y.; Tsoi, L.C. A Review of Recent Advancement in Integrating Omics Data with Literature Mining towards Biomedical Discoveries. *Int. J. Genom.* **2017**, *2017*, 6213474. [[CrossRef](#)] [[PubMed](#)]
86. Ritchie, M.D.; Holzinger, E.R.; Li, R.; Pendergrass, S.A.; Kim, D. Methods of integrating data to uncover genotype-phenotype interactions. *Nat. Rev. Genet.* **2015**, *16*, 85–97. [[CrossRef](#)] [[PubMed](#)]
87. Yan, S.K.; Liu, R.H.; Jin, H.Z.; Liu, X.R.; Ye, J.; Shan, L.; Zhang, W.D. “Omics” in pharmaceutical research: Overview, applications, challenges, and future perspectives. *Chin. J. Nat. Med.* **2015**, *13*, 3–21. [[PubMed](#)]
88. Kim, T.Y.; Kim, H.U.; Lee, S.Y. Data integration and analysis of biological networks. *Curr. Opin. Biotechnol.* **2010**, *21*, 78–84. [[CrossRef](#)] [[PubMed](#)]
89. Civelek, M.; Lusis, A.J. Systems genetics approaches to understand complex traits. *Nat. Rev. Genet.* **2014**, *15*, 34–48. [[CrossRef](#)] [[PubMed](#)]
90. Yugi, K.; Kubota, H.; Hatano, A.; Kuroda, S. Trans-Omics: How to Reconstruct Biochemical Networks across Multiple ‘Omic’ Layers. *Trends Biotechnol.* **2016**, *34*, 276–290. [[CrossRef](#)] [[PubMed](#)]
91. Sun, Y.V.; Hu, Y.J. Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases. *Adv. Genet.* **2016**, *93*, 147–190. [[PubMed](#)]
92. Meng, C.; Zeleznik, O.A.; Thallinger, G.G.; Kuster, B.; Gholami, A.M.; Culhane, A.C. Dimension reduction techniques for the integrative analysis of multi-omics data. *Brief. Bioinform.* **2016**, *17*, 628–641. [[CrossRef](#)] [[PubMed](#)]
93. Huang, S.; Chaudhary, K.; Garmire, L.X. More Is Better: Recent Progress in Multi-Omics Data Integration Methods. *Front. Genet.* **2017**, *8*, 84. [[CrossRef](#)] [[PubMed](#)]
94. Lin, E.; Lane, H.Y. Machine learning and systems genomics approaches for multi-omics data. *Biomark. Res.* **2017**, *5*, 2. [[CrossRef](#)] [[PubMed](#)]
95. Morota, G.; Ventura, R.V.; Silva, F.F.; Koyama, M.; Fernando, S.C. Machine learning and data mining advance predictive big data analysis in precision animal agriculture. *J. Anim. Sci.* **2018**. [[CrossRef](#)] [[PubMed](#)]
96. Luo, G.; Stone, B.L.; Johnson, M.D.; Tarczy-Hornoch, P.; Wilcox, A.B.; Mooney, S.D.; Sheng, X.; Haug, P.J.; Nkoy, F.L. Automating Construction of Machine Learning Models with Clinical Big Data: Proposal Rationale and Methods. *JMIR Res. Protoc.* **2017**, *6*, e175. [[CrossRef](#)] [[PubMed](#)]
97. Remington, R.W.; Yuen, H.W.; Pashler, H. With practice, keyboard shortcuts become faster than menu selection: A crossover interaction. *J. Exp. Psychol. Appl.* **2016**, *22*, 95–106. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).