Chapter 11
Management of Next-Generation Sequencing in Precision Medicine

Shing Cheng Tan, Hui-min Neoh, Mia Yang Ang, Mohamad Ayub Khan Sharzehan, Nursyazwani Omar, and Teck Yew Low

Abstract Next-generation sequencing (NGS) has transformed DNA sequencing in terms of speed and data volume, rendering genomics affordable and achievable by individual laboratories rather than big science that was once managed by international consortiums. In parallel, it is propelling contemporary healthcare into the age of precision medicine, whereby genetic variability of an individual is incorporated into the formulation of a bespoke treatment plan. However, due to the complexity of NGS workflow and the large volume of both NGS and phenotypic data, management of the NGS "big data" in precision medicine has been challenging. This chapter mainly discusses these challenges and their solutions from several perspectives, including (i) sample logistics; (ii) electronic health records; (iii) sequencing procedures; (iv) bioinformatics analysis; (v) interpretation and delivery of results; (vi) the storage and reanalysis of NGS data; and (vii) the laboratory information management system (LIMS) as an overall management suite. Finally, we outline several current developments including artificial intelligence (AI), Internet of Things (IoT), wearable technologies, and 5G communication. Presumably, these technologies are capable of bringing impacts to precision medicine in terms of the range of phenotypic data, which can be acquired continuously and transferred with enhanced connectivity and speed and in real time.

Keywords Big data · Data management · LIMS · Personalized medicine

S. C. Tan · H.-m. Neoh · M. A. K. Sharzehan · N. Omar · T. Y. Low (✉)
UKM Medical Molecular Biology Institute, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia
e-mail: lowteckyew@ppukm.ukm.edu.my

M. Y. Ang
Department of Gene Diagnostics and Therapeutics, National Center for Global Health and Medicine, Tokyo, Japan

Department of Clinical Genome Informatics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

H. Shen et al. (eds.), *Regionalized Management of Medicine*, Translational Bioinformatics 17, https://doi.org/10.1007/978-981-16-7893-6_11

11.1 Introduction

Precision medicine (PM) is an emerging paradigm in contemporary healthcare. In comparison to how medicine is currently practiced, PM takes into account the variability of an individual such as one's genetics, lifestyle, and environment prior to formulating and customizing a treatment plan so as to provide a precisely tailor-made therapy to each patient. Consequently, PM allows accurate predictions of disease susceptibility, diagnosis, and prognosis and has been touted as the only way forward for optimizing patient care (Roberts and Julius 2016). One major driving force behind PM is next-generation sequencing (NGS) which offers multiple advantages over the past generations of DNA sequencing technologies (Morash et al. 2018; Gonzalez-Garay 2014). For example, NGS (i) is highly sensitive; (ii) allows simultaneous sequencing of millions of genetic fragments at a high coverage and depth; (iii) requires only a low input of starting material; and (iv) is capable of detecting chromosomal aberrations, copy number variations, and novel low-frequency genetic variants.

Owing to the strength of NGS, several high-profile genomic initiatives have employed NGS for cataloging molecular variations in human diseases (Table 11.1). One example is The Cancer Genome Atlas (TCGA), which has profiled the molecular landscape of 33 tumor types derived from over 11,000 patients (Blum et al. 2018). TCGA data have contributed meaningfully to precision oncology. For instance, based on the molecular profiles, gliomas can now be classified into distinct subtypes, namely, those with IDH/t(1p;19q)/TERT alterations; IDH/TERT mutations; IDH mutations only; TERT mutations only; or triple-negative gliomas (Eckel-Passow et al. 2015). Similarly, gastric cancer can be categorized into four or five different subtypes (depending on the classification system used), including a microsatellite unstable subtype and an Epstein-Barr virus (EBV)-positive subtype (Bass et al. 2014; Setia et al. 2016). Besides tumor classification, NGS data generated from TCGA also revealed several actionable mutations and gene fusions in several subtypes of cancer, which may provide insights for targeted therapies (Gao et al. 2018; Grabiner et al. 2014; Wagle et al. 2014).

Another PM initiative enabled by NGS is the Human Microbiome Project (HMP). The first phase of the HMP established the baseline taxonomic and strain-specific compositions of the microbiome within and between body sites, whereas the second phase characterized the dynamic changes in microbiome and host profiles during pregnancy and preterm birth; the onset of inflammatory bowel diseases; and the onset of type 2 diabetes (Sinha et al. 2015; The Integrative Human Microbiome Project 2019; Proctor et al. 2019; The Integrative HMP (iHMP) Research Network Consortium 2014). HMP data improve our understanding of host-microbiome interactions and their underlying mechanisms. As the roles of host-microbiome interactions become more apparent, NGS data from the HMP can be tapped to facilitate the development of precision diagnostics and therapeutics (ElRakaiby et al. 2014; Doestzada et al. 2018).

Table 11.1 Examples of high-profile genomic initiatives which employed NGS for molecular profiling

Initiative	Launch year	Goal
The Cancer Genome Atlas (TCGA)	2005	To catalog and discover major cancer-causing genomic alterations through large-scale genome sequencing and integrated multi-dimensional analyses
Human Microbiome Project	2007	To characterize the human microbiome and analyze its role in human health and disease
Encyclopedia of DNA Elements Consortium (ENCODE)	2003	To identify all functional elements in the human and mouse genomes, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active
1000 Genomes Project	2008	To identify genetic variants with frequencies of at least 1% in the populations studied
100,000 Genomes Project	2012	To sequence 100,000 genomes from around 85,000 UK National Health Service (NHS) patients affected by rare diseases, cancers, and infectious diseases
1+ Million Genomes Initiative	2018	To sequence at least one million genomes in the European Union for improving disease prevention, allowing for more personalized treatments, and providing a sufficient scale for new clinically impactful research
All of Us Research Program	2015 (enrollment started 2018)	To collect genetic and health data from one million volunteers to accelerate health research and medical breakthroughs, enabling individualized prevention, treatment, and care

Several other large-scale projects have also exploited NGS for decoding the genome, epigenome, transcriptome, regulome, and interactome of the study participants. These efforts include the Encyclopedia of DNA Elements Consortium (ENCODE) project (Dunham et al. 2012), the 1000 Genomes Project (Auton et al. 2015; Sudmant et al. 2015), and more recently the 100,000 Genomes Project of the United Kingdom (Peplow 2016; Turnbull et al. 2018), the 1+ Million Genomes Initiative of the Europe (Saunders et al. 2019), and the All of Us Research Program of the United States (US) (Sankar and Parker 2017). Many smaller-scale studies also utilized NGS for identifying molecular biomarkers that can improve precision diagnostics and prognostication (Alonso et al. 2019; Zacher et al. 2017; Acha et al. 2019; Na et al. 2019; Oberg et al. 2016; Volckmar et al. 2019; Dubbink et al. 2016). The widespread use of NGS in these projects highlights its potential in accelerating PM. Nevertheless, the large volume of data generated by NGS poses a challenge in its management. In this chapter, we first describe the general workflow of NGS and its challenges and further discuss the use of laboratory information management system (LIMS) in managing NGS.

11.2 Workflow of Next-Generation Sequencing Experiments

Genome-scale PM projects that incorporate NGS workflows are often complicated, consisting of multiple steps that require multidisciplinary expertise. Due to such complexity, the success of such projects relies heavily on excellent management and logistics of samples, NGS data, patients' records, and other ancillary data. In the following, we discuss the current practice of NGS data management, paying particular attention to the following aspects: (i) the management of sample logistics; (ii) archiving and retrieval of patients' electronic medical records; (iii) the sequencing procedures; (iv) bioinformatics analysis; (v) the interpretation and delivery of results; (vi) the storage and reanalysis of archived NGS data; and (vii) the laboratory information management system (LIMS) to enable seamless and overall connection and coordination for NGS samples, workflows, and data.

11.2.1 Sample Acquisition and Management

Managing sample logistics is the very first step of NGS-based projects. Once a patient has provided informed consent to a PM regimen, the next crucial step is to ensure acquisition of high-quality sample(s) from patients. To do this, there must be stringent quality control in the collection, temporary and long-term storage, as well as processing of samples (Moore et al. 2011). As such, prior to sample collection, patients should have undergone consultation and counseling by clinicians and genetic counselors; have agreed to various terms and conditions of the diagnosis/treatment workflow; and have been briefed about the sample collection process. While all patient- and sample-related information should be documented, personal identifiers should strictly be kept confidential. Depending on the nature of investigations, biospecimens, such as blood, saliva, buccal swabs, urine, and stool, will be acquired from the patients (Basik et al. 2013; Woo and Lu 2019; van Noord 2003; Malsagova et al. 2020). Drawing of blood requires the assistance of a qualified phlebotomist, while acquisition of buccal swabs, urine, and stool may be performed by the patients themselves. Some tissues, such as tumor tissues or endoscopic swabs, can only be obtained with the help of clinicians in hospitals. Care must be taken by all process operators to minimize the contamination of samples with other sources of nucleic acid.

Once obtained, biospecimens should be kept at temperatures optimal for nucleic acid extraction. A transport temperature of 4 °C is deemed optimal for delivery of samples for nucleic acid extraction; nevertheless, delivery at 25 °C or below for samples such as buccal swabs may be accepted, if cold chain is not available. Nucleic acids extracted from the biospecimens should be kept at −20 °C until NGS is performed. Indeed, best practices and standardization for quality control of biospecimens prior to analytical process have been developed by various

organizations, including the Clinical and Laboratory Standards Institute (CLSI) Standards, the National Cancer Institute (NCI) Best Practices (NCI Best Practices for Biospecimen Resources 2016), and the College of American Pathologists (CAP) Guidelines (Stumptner et al. 2019). In addition, laboratories involved in the NGS workflow should ideally be accredited with ISO 15189:2012 (Medical Laboratories – Requirements for Quality and Competence) (Gutowska-Ding et al. 2020).

Biobanking of samples plays an important role in PM research. Therefore, patients may choose to bank their samples or nucleic acids for follow-up/future investigation. It is also imperative that biobanks adopt clearly defined guidelines on the ownership, usage, and disposal or destruction of patient samples to ensure high quality of archived samples for future usage (Malsagova et al. 2020; Ollier et al. 2005; Nagai et al. 2017; Langhof et al. 2018).

11.2.2 Medical and Health Informatics

The Health Information System (HIS) is an indispensable component of contemporary healthcare management. Notably, many healthcare providers have implemented electronic health records (EHRs) for sharing and managing clinical data according to predetermined standards and classifications (Kho et al. 2013). Such data categories encompass billing information, time of visitation, clinical symptoms and interpretations, prescription, laboratory test results, and other patient-linked data (Jensen et al. 2012). Hence, EHRs customarily adopt a structured format, comprising unchanged administrative data such as demographic details and updatable particulars such as disease diagnoses and treatment procedures (Wu et al. 2017). EHRs may also contain unstructured data. For instance, patients' clinical conditions can be described in free-text clinical notes to supplement existing structured data; and therefore natural language processing (NLP) capability has been incorporated in newer EHR systems (Nadkarni et al. 2011). A non-exhaustive list of EHR software in the market and their features is shown in Table 11.2.

In 2015, the US Precision Medicine Initiative proposed to incorporate omics data into EHRs (Wu et al. 2017; Collins and Varmus 2015). Such a link extends the roles of EHRs to medical research, whereby clinical information and healthcare-related biological products can be used to supplement genetic investigation of diseases at both individual and population-wide levels and to drive large cohort studies (Kho et al. 2013; Kohane 2011). Vice versa, omics modalities, such as NGS data, together with EHRs can be exploited for the study of healthcare efficiency and safety across the population (Wu et al. 2017). Technically, EHR-driven genomic research (EDGR) is an emerging area that employs a combination of structured, codified, and narrative texts to describe phenotyping, sample selection, and acquisition. Alternatively, clinical characterization can be applied as an addition to fill in missing details of collected samples as a form of phenotypic augmentation (Kohane 2011). Indeed, NGS data has assisted the development of drugs targeting tumor-driving mutations (Morash et al. 2018). Besides, by linking biorepositories to clinical data,

Table 11.2 A list of electronic health records (EHR) software in the market and their features

No.	Vendor name	Product name	Features
1.	Varian Medical Systems	ARIA® Oncology Information System	• Cloud-based and on premise • Support clinical workflow and document management • No insurance and claim information • No lab integration • Demographics and portal of patients • Reporting and analytics
2.	Elation Health	Elation HER	• Cloud-based only • Support clinical workflow and document management • No insurance and claim information • Lab integration • Demographics, history, portal, and referrals of patients • Reporting and analytics
3.	Allscripts	TouchWorks HER	• Cloud-based only • Support clinical workflow and document management • Contains insurance and claim information • Lab integration • Demographics, history, portal, and referrals of patients • Reporting and analytics
4.	Epic	Epic HER	• Cloud-based only • Support clinical workflow and document management • Contains insurance and claim information • Lab integration • Demographics, history, portal, and referrals of patients • Reporting and analytics
5.	eClinicalWorks	eClinicalWorks 10e	• Cloud-based only • Support clinical workflow and document management • Contains insurance and claim information • Lab integration • Demographics, portal, and referrals of patients • Reporting and analytics
6.	GE Healthcare	Centricity EMR	• Cloud-based only • Support clinical workflow and but not document management • No insurance and claim information • Lab integration • Demographics and history of

(continued)

Table 11.2 (continued)

No.	Vendor name	Product name	Features
			patients • Reporting and analytics
7.	Flatiron	OncologyCloud	• Cloud-based only • Support clinical workflow and document management • No insurance and claim information • No lab integration • Demographics, history, and portal of patients • Reporting and analytics
8.	ChartLogic	ChartLogic EHR	• Cloud-based and on premise • Support clinical workflow and document management • Contains insurance and claim information • Lab integration • Demographics, history, portal, and referrals of patients • Reporting and analytics
9.	Bizmatics	PrognoCIS	• Cloud-based only • Support clinical workflow and document management • Contains insurance and claim information • Lab integration • Demographics, history, portal, and referrals of patients • Reporting and analytics
10.	AdvancedMD	AdvancedEHR	• Cloud-based only • Support clinical workflow and document management • Contains insurance and claim information • Lab integration • Demographics, history, portal, and referrals of patients • Reporting and analytics
11.	CureMD	All in One EHR	• Cloud-based only • Support clinical workflow and document management • Contains insurance and claim information • Lab integration • Demographics, history, portal, and referrals of patients • Reporting and analytics
12.	RXNT	RXNT EHR	• Cloud-based only • Support clinical workflow and document management

(continued)

Table 11.2 (continued)

No.	Vendor name	Product name	Features
			• Contains insurance and claim information • Lab integration • Demographics, history, portal, and referrals of patients • Reporting and analytics
13.	MDVision	MDVision EMR	• Cloud-based and on premise • Support clinical workflow and document management • Contains insurance and claim information • Lab integration • Demographics, history, portal, and referrals of patients • Reporting and analytics
14.	Kareo	Kareo Clinical	• Cloud-based only • Support clinical workflow and document management • Contains insurance and claim information • Lab integration • Demographics, history, portal, and referrals of patients • Reporting and analytics

more sophisticated and personalized disease management can be followed through by clinician-scientists. Furthermore, the information on family history and genetic testing can also help to support and educate clinicians in adopting evidence-based strategies in clinical care and decision support (Scheuner et al. 2009). By referring to sequencing results to identify a patient's cancer genome, for instance, both treatment options and management can be facilitated; and this has yielded benefits in providing assessment of new treatments, improvements in patient outcomes, or healthcare savings (Cowie et al. 2017).

While transparent health information systems facilitate data sharing and collaborative research, it is noteworthy that impending risks may arise in the form of infringements of privacy, social stratification and discrimination, or identity theft. This can lead to the erosion of public trust in healthcare providers and governments. Therefore, management of EHRs should preferably incorporate a legislative framework of privacy protection, besides issues in data security and cyber-security.

11.2.3 Next-Generation Sequencing Techniques

Selecting appropriate NGS technology platforms, the associating data structures and formats, as well as the bioinformatics pipelines are important determinants for downstream data management such as storage and retrieval. The first commercial NGS technology was initially launched in 2004 by the 454 Life Sciences based on the pyrosequencing method (Margulies et al. 2005). Subsequently, Illumina launched its "sequencing by synthesis" technology—the principle behind Illumina's fleet of sequencers such as Solexa, MiSeq, HiSeq, and NextSeq (Meyer and Kircher 2010). Life Technologies later launched the SOLiD platform (Valouev et al. 2008), which introduced "sequencing by ligation," and, later, the Ion Torrent and Ion Proton semiconductor sequencing platforms, which perform "sequencing by synthesis" (Pennisi 2010). Following this, Beijing Genome Institute (BGI) acquired Complete Genomics's "DNA nanoballs" sequencing by ligation technology with its BGISEQ-500 flagship sequencer (Porreca 2010). In 2013, the 454 pyrosequencing platform was discontinued by Roche, while Thermo Fisher Scientific, which acquired Life Technologies, discontinued SOLiD sequencing in 2016. This leaves Illumina, Ion, and BGI sequencing as the main players in NGS technologies. The timeline for major development milestones of both NGS and TGS platforms is illustrated in Fig. 11.1.

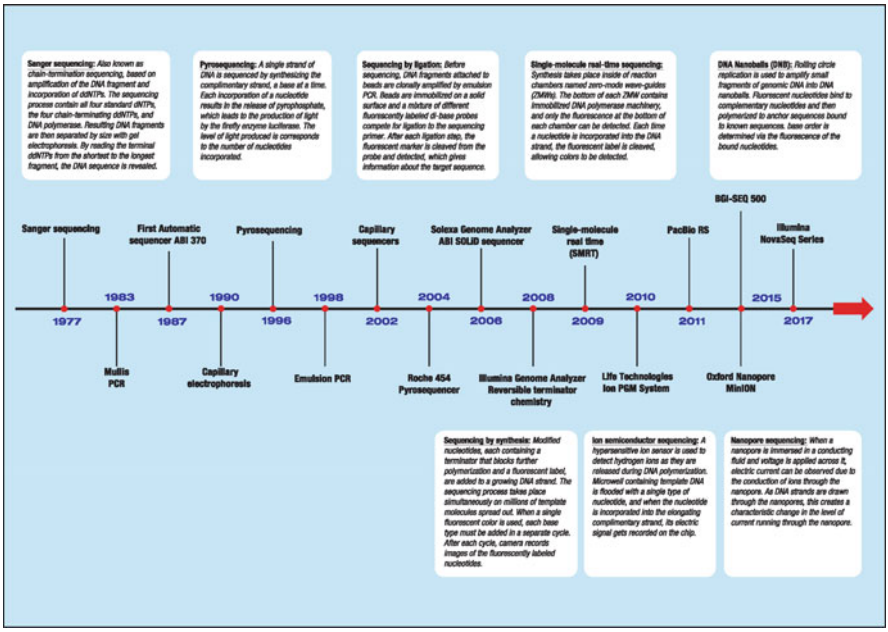

Fig. 11.1 Timeline for major development milestones of both NGS and TGS platforms

Almost all NGS platforms produce short sequencing reads of between 100 and 400 bases. Short sequences can sometimes be difficult to assemble (Margulies et al. 2005). In contrast, PacBio and Oxford Nanopore introduced long-read sequencing, where reads from these third-generation sequencing (TGS) platforms can span 10 kb or longer. TGS reads are easier to assemble, and they may reveal structural variation and allow easier determination of maternal and paternal chromosomes for neonatal PM (Rhoads and Au 2015; Tyson et al. 2017).

Regardless of NGS or TGS, the sequencing process can be divided into four phases, namely, library preparation, amplification, sequencing, and data analysis. Besides whole genome sequencing (WGS) in cases of unknown causes of disease, patients can opt for the smaller whole exome sequencing (WES) if suspected mutations or single-nucleotide polymorphism (SNPs) of genes of interest are confirmed to be located in the exons. Sequencing of targeted-gene panels will reveal mutations or SNPs which have been clinically proven to be related with specific diseases or treatment outcomes. In cases where diseases or traits are confirmed or suspected to be epigenetically related, patients may also opt for epigenome and transcriptome sequencing (RNA-Seq) (Maróti et al. 2018; Zhang et al. 2018). Besides, as the role of specific microbes in health and disease becomes more apparent, microbiome profiling may soon find its immediate applicability in PM (ElRakaiby et al. 2014; Doestzada et al. 2018).

11.2.4 Bioinformatics

NGS technology generates a plethora of data. These data need to be bioinformatically analyzed before they can be translated into clinically meaningful information. The general bioinformatics analysis workflow for NGS is shown in Fig. 11.2. The first-hand data that have not been modified since acquisition are termed raw data. The file types and structures of raw data vary according to the platforms used, but they are generally large in size. Raw data typically contains noises from upstream analyses. Removal of these noises results in processed data, which can be further analyzed to obtain clinically meaningful information. Nonetheless, processed raw data should preferably be stored to allow reanalysis with future versions of data processing algorithms which can achieve greater precision and accuracy (Hart et al. 2016). In addition to raw data, cryptographic hash such as SHA or MD5 should be generated and stored to prevent data corruption. Also, data stored on local computers or institutional servers should be backed up periodically to other locations to protect against data loss (Mangul et al. 2018).

NGS data are often flooded with sequencing artifacts, which include reads error, poor-quality reads, and primer or adaptor contamination (Endrullat et al. 2016). Hence, quality control (QC) of NGS data is vital to filter out low-quality data that would have imposed negative impacts on the downstream analyses and misleading conclusions (Patel et al. 2012). Several different software programs can be used for converting sequence data into nucleotide reads and their annotation. These programs

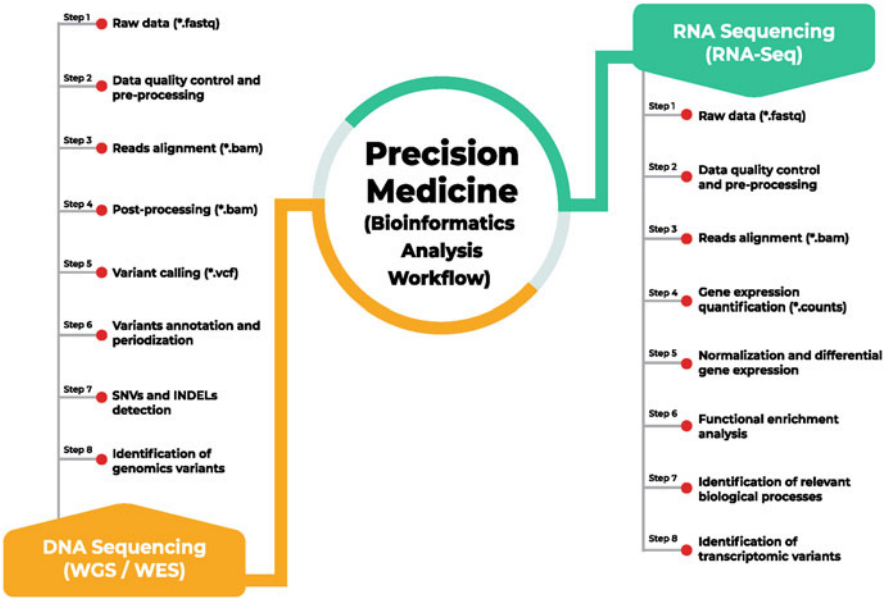

Fig. 11.2 A step-by-step bioinformatics analysis workflow for DNA-Seq and RNA-Seq acquired with NGS or TGS platforms

can be broadly classified into (i) open-source software and (ii) commercial software. While open-source programs are freely available and allow user preview of the logical flow and customization of the pipelines, programming skills are needed to execute and implement the program (Mangul et al. 2018). Commercial software, on the other hand, combines several analytical programs in a package and is usually equipped with a point-and-click interface. The downside of commercial software is that they are expensive in terms of licensing and upgrading costs (Smith 2014). Both open-source and commercial software receive version updates from their developers regularly. While the new updates are usually hassle-free for commercial software (since they are all contained in a suite package), open-source software requires additional checking for hardware and software compatibility issues and might need to recompile from codes again, which makes the time-stamping and management of the software versions more complicated (Gullapalli 2020).

11.2.5 Delivery of Results

To apply NGS data to the clinical setting, NGS data need to be translated into outputs that are informative to the healthcare providers as well as to the patients (Manrai and Kohane 2017; Mehandziska et al. 2020). To this end, bioinformaticians play an important role in identifying changes in the tested samples in comparison to controls.

Depending on the tests, bioinformaticians and geneticists/molecular biologists will need to work together to match these changes to available literature and databases on the investigated illness/traits and deliver the results to medical practitioners for patient disclosure. In cases where the cause of the investigated disease/trait is still unclear, a multidisciplinary team composed of bioinformaticians, scientists, and clinicians will be important for NGS results interpretation (Bylstra et al. 2019).

During results disclosure to individuals carrying high genetic risks, it is important to have medical professionals explaining the biological meaning of the results. It will also be ideal to have genetic counseling at the same time or at a subsequent session (Stoll et al. 2018). This is particularly important if the investigated condition is hereditary, so that the patients are made aware of not only their health condition but also the probable impact on their families and offsprings. Subsequently, the tested patient and/or probands can go through a confirmatory testing or be referred to specialist clinics for clinical management (Bylstra et al. 2019). On the other hand, if the patient is not found to carry any pathogenic mutation, they can be informed via an official letter from the clinic, with the option of a final meeting with the attending clinician, if required.

As NGS-based PM results contain sensitive information about the tested patient's genome, all folders containing these data should be encrypted with no personal identifiers to patients (Martinez-Martin and Magnus 2019). After consultation and counseling, it depends on the patient's prior discussion and agreement with the attending clinician on how their data will be managed after NGS results are delivered – whether this data should be destroyed or archived for future usage.

11.2.6 Data Storage

Storage of NGS data allows data reanalysis in the future. With the surge of NGS datasets, the demand in computing power and storage has increased exponentially. The processing of large-scale NGS data often requires high-performance computing (HPC) resources (Pérez-Wohlfeil et al. 2018). Despite enhancing productivity, HPC systems harbor security risks due to possible attacks by hackers, an inherited problem from conventional computer, and network security issues (Hou et al. 2020). Since HPC systems are essentially large-scale computing infrastructure, they can be remotely accessed by any registered users. Therefore, security mechanisms must be employed to prevent any suspicious activities or attempts from accessing sensitive information.

The increase in the volume and variety of NGS data has posed new challenges, where HPC is now dealing with big data garnering terabytes/petabytes of information. By tying together all aspects of computing power and storage, cluster computing helps to eliminate gridlocks that can interfere with performances (Ocaña and De Oliveira 2015). The greatest advantage of computer clusters is the scalability that they offer, in which, unlike mainframe computers that have a fixed processing

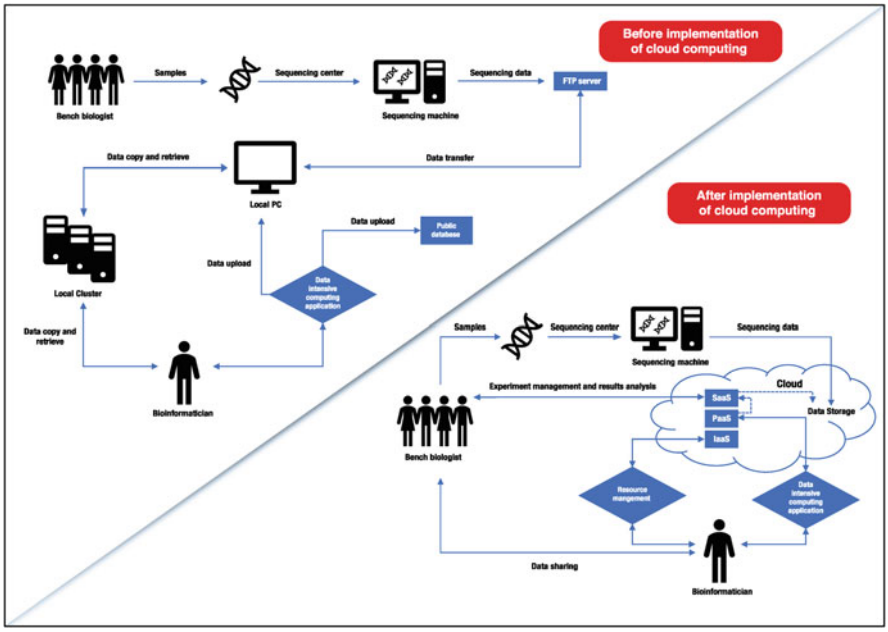

Fig. 11.3 Computational infrastructure, requirements, and logistics for NGS data analysis – before and after implementation of cloud computing

capacity, computer clusters can be easily expanded as requirements change and grow, simply by adding additional nodes and storages to the cluster.

Bioinformatics services have generally been provided via web servers, and are usually hosted at institutional computing infrastructures, and simultaneously serve multiple users via remote access. Due to the expanding number of users, data sizes, and new requirements in terms of speed and availability, this model has become outdated (El-Kalioby et al. 2012). In recent years, cloud computing and storage have emerged as a powerful, flexible, and scalable approach to tackle computational and data-intensive problems (Fig. 11.3). Examples of public clouds include Amazon Web Services (AWS), Google Cloud Platform (GCP), and Microsoft Azure. Several common cloud types include (i) software as a service (SaaS), which enables the user to use the cloud provider's applications that are running on the cloud infrastructure; (ii) platform as a service (PaaS), which enables the user to create or acquire applications and tools and deploy them on the cloud provider's infrastructure; and (iii) infrastructure as a service (IaaS), which enables the user to utilize processing, storage, networks, and other fundamental computing resources (Navale and Bourne 2018).

11.3 Challenges in Next-Generation Sequencing

Apart from data storage, there are several other technical challenges in applying NGS in precision medicine. First, there is a need for quality management and workflow standardization (Endrullat et al. 2016). This ensures the comparability of the sequencing data among different laboratories and reproducibility of the results, thus improving testing accuracy and reliability. Currently, there is a lack of consensus in the development of a standardized workflow among different laboratories. Even when standardization guidelines are developed, there might be difficulties implementing the standards due to incompatibility of the guidelines in different NGS applications (Cargill 2011). Moreover, a vast majority of standardization guidelines are from the United States; thus, it is unknown whether these guidelines are implementable in other countries (Endrullat et al. 2016). Apart from that, there is also an issue concerning the bandwidth limitations if sequencing and data analysis are performed at different centers, thus requiring data transfer from one place to another (Wandelt et al. 2012). In addition, as NGS involves multiple procedures over multiple days which may be performed by multiple technologists, it is often difficult to track the NGS workflow and trace the samples, the personnel who performed the experiment, reagent consumption, etc.

11.4 Laboratory Information Management System

Fortuitously, laboratory information management systems (LIMS) are available, which can be used to manage NGS data and workflow, and make the tracking and tracing above uncomplicated. LIMS is a type of software developed and used in many clinical laboratories since the 1980s to manage basic operations in the laboratory. Traditionally, the functions of LIMS focused on the management of samples, such as sample registration, barcode label generation, and location tracking (including assigning the samples to a particular freezer compartment) (Gibbon 1996). Over time, the functionalities have been enhanced by including management of personnel (such as assignment of job to laboratory staff and their work schedules); estimation and tracking of experiment completion time; assessment of instrument or test performance; reagent and inventory tracking; recording of experimental procedures (i.e., electronic laboratory notebook); and tax invoicing (Argento 2020; De Block 2019). Implementation of LIMS in clinical laboratories can reduce human errors during the entire experimental workflow and aid in the enhancement of work quality and productivity and is among the first steps toward laboratory automation and digitalization (Chen et al. 2016; Junaid and Jangda 2020).

Despite this, the unprecedented volume of NGS data, along with the complexity of the NGS workflow, poses new challenges to the conventional LIMS (Chen et al. 2016). For this reason, many conventional LIMS have been refined or even revamped, and many other new LIMS have been developed, to meet the demands

of modern genomic laboratories. For example, apart from tracking the identity of the samples and their location in the freezer, modern LIMS also records the status of the samples (e.g., whether they are successfully completed) at each stage of the complex NGS pipeline, as well as the batch number of the reagents used. Modern LIMS also allows protocol automation and customization for different NGS applications and platforms and contains quality control tools that identify samples of poor quality. Besides, many modern LIMS have raw data processing functionalities, which can ease the delivery of sequencing results, often in a systematic and organized manner, to the healthcare providers or the patients. Due to the enormous size of NGS data files, modern LIMS are also moving toward cloud computing and storage, as discussed in Sect. 11.2.6 (Paul et al. 2017; Guo et al. 2020). For this reason, most modern LIMS applications require connectivity with a high-speed internet and normally function through a server computer and several PC workstations.

11.5 How to Select a Laboratory Information Management System?

Dozens of LIMS offering different functionalities are currently available in the market, and each one has its own pros and cons. Several main aspects that need to be considered when selecting a LIMS for NGS management are discussed below.

11.5.1 Ability to Integrate with NGS Instrumentation

To improve the workflow efficiency in the laboratory, a LIMS needs to be able to directly provide instructions or data to, or retrieve data from, the existing NGS instruments. In other words, a LIMS should be integrated with NGS instrumentations rather than operating independently (Chen et al. 2016). This will allow automation of the process, instead of relying on laboratory staff to manually feed or retrieve the information between LIMS and the instruments. It should be noted that distinct NGS platforms use different reagents, protocols, and instruments to achieve their optimal performance. Ideally, a LIMS should be able to recognize and communicate with instrumentations from all major platforms, so that users can simply specify the type of NGS application to be performed and the LIMS automatically generates the correct instructions for the sequencers. In addition, the ability to integrate with NGS instrumentations can allow LIMS to continuously monitor metrics such as the Phred quality score to track the status and quality of the sequencing in real time.

11.5.2 Ease of Configuration and Customization

Every laboratory has its own demands for LIMS. In addition, NGS workflows, protocols, methodologies, and technologies are also continuously evolving (Durmaz et al. 2015; Cała et al. 2014). Thus, it is imperative that LIMS be flexible for configuration and customization (Chen et al. 2016; Tagger 2011). Configuration of LIMS refers to the use of built-in system tools, often in the user interface, to redesign the appearance or functionality of the software. Examples of configuration tasks include creating a new user profile, adding new sample preparation procedures or analytical methodologies, devising the sample nomenclature, and inserting additional fields in the report form. Customization, on the other hand, refers to extension of the system functionalities by modifying the programming codes associated with the software. This is necessary when, for example, the LIMS needs to be catered for new NGS technologies or instrumentations. Customization often requires the skills of expert programmers and involves rigorous testing and validation of the new codes to ensure that they do not clash with the system (Paul et al. 2017; Bianchi et al. 2016; Rafid Raeen 2018). Thus, it is clear that LIMS which allows easy configurations and customizations can allow laboratories to work efficiently to match their current and future informatics needs.

11.5.3 Accommodativeness to Different Users

Implementation of NGS in the clinical practice typically requires the involvement of several parties including laboratory managers, technicians, application specialists, bioinformaticians, authorized signatories, and clinicians/healthcare providers. Each of these parties has their own responsibilities and thus requires access to different information (Tagger 2011). For example, a technician does not need to know the medical history of a patient, whereas a healthcare provider does not need to know the rate of reagent consumption for an NGS assay. To improve work efficiency, a LIMS should provide targeted interfaces for different users, so that each of them can access only the information that is relevant to them, rather than spending time retrieving the necessary data from the system.

11.5.4 Comprehensiveness of Sample Tracking

The accuracy of an NGS assay is dependent on the quality of the sample in each step of the workflow ("garbage in, garbage out") (Robles-Espinoza and Adams 2014; Conrads and Abdelbary 2019). As such, it is important to keep close track of the samples throughout the entire workflow, from sample collection to final result delivery (Matthijs et al. 2016). To enable effective analysis of a large volume of

NGS data, LIMS needs to be able to track the samples and their associated metadata comprehensively. Examples of information that needs to be captured by a LIMS include the approach by which the samples were collected, the purity and concentration of the nucleic acids isolated, sample and nucleic acid storage conditions, the protocols and reagents (including batch numbers) used to create the sequencing library, quality control metrics of the sequencing library, etc. A comprehensive sample tracking can allow validation and/or troubleshooting work to be performed effectively.

11.5.5 Compliance Support

The use of NGS in precision medicine involves human samples and is thus governed by a number of regulations. In the United States, for example, healthcare providers need to adhere to the Health Insurance Portability and Accountability Act (HIPAA), Clinical Laboratory Improvement Amendments (CLIA), and General Data Protection Regulation (GDPR), in addition to other guidelines and requirements from international agencies and professional organizations (see Sect. 11.2.1) (Roy et al. 2018). Apart from that, an NGS laboratory should ideally comply with the ISO 15189:2012 accreditation standards (Gutowska-Ding et al. 2020). A good LIMS should have a set of controls that monitor compliance with these regulations, guidelines, and standards to ensure integrity and validity of the NGS data and to safeguard patients' privacy.

11.5.6 Cost

Both commercial and non-commercial (open-source) LIMS are widely available in the market. Commercial LIMS are usually easier to install and use, and have more advanced features, compared to the open-source ones. However, the costs needed to implement a commercial LIMS are often unaffordable for smaller laboratories which do not run NGS as the core business and rely on research grants as the main source of funding (Lemmon et al. 2011).

11.6 Examples of LIMS for NGS

The rise of NGS has also pushed the rapid development of LIMS. A number of LIMS are now available to manage NGS data (Table 11.3). Laboratories using the Illumina NGS platform are probably familiar with the Illumina BaseSpace Clarity (BaseSpace Clarity LIMS 2020), a web-based LIMS which can receive encrypted data directly from the sequencers and used for monitoring sequencing runs,

Table 11.3 A non-exhaustive list of commercial and open-source LIMS for NGS

LIMS	Licensing
Illumina BaseSpace Clarity	Commercial
Agilent SLIMS	Commercial
Exemplar	Commercial
GNomEx	Open-source[a]
openBIS	Open-source
Wasp	Open-source
MendeLIMS	Open-source
SLIMS	Open-source
SMITH	Open-source
Galaxy LIMS	Open-source
Parkour LIMS	Open-source
SeqBench	Open-source
MISO	Open-source
NG6	Open-source

[a]Free disk space allocation is 100 GB. A bill of $2.00 will be charged each month for every additional 100 GB ($240 per TB per year)

processing and storing the sequencing data on the BaseSpace Cloud, and easy sharing of the sequencing data with collaborators. Likewise, Agilent users may be familiar with SLIMS, which provides comprehensive tools for managing the entire NGS workflow. Another LIMS which can be used for NGS data management is openBIS (Barillari et al. 2016), which also features an electronic laboratory notebook function. GNomEx LIMS (GNomEx 2020), which is developed by the Huntsman Cancer Institute, can also be used for reporting of both NGS and microarray data. The Wasp system (McLellan et al. 2012) is another genomics LIMS that provides embedded pipelines for various massively parallel sequencing assays for use in the clinic. Similarly, MendeLIMS (Grimes and Ji 2014) also emphasizes on the management of clinical genome sequencing projects. The Sample-based Laboratory Information Management System (SLIMS, not to be confused with the Agilent SLIMS which shares the same name above) (Van Rossum et al. 2010), on the other hand, is a LIMS designed for laboratories focusing on genotyping and genome-wide association studies. In addition, SMITH (Venco et al. 2014) provides a fully automated system in managing sequencing data on high-performance computing clusters. Besides, the Galaxy LIMS (Scholtalbers et al. 2013) was developed as an extension to the existing Galaxy platform; thus, the sequencing data are directly available for processing using the analytical pipelines stored in the platform. The Parkour LIMS (Anatskiy et al. 2019) is designed to maximize the efficiency of NGS sample processing and quality management in academic core laboratories. SeqBench (Dander et al. 2014) is another LIMS that is developed to cater analysis of whole exome sequencing data, either locally, on a cluster, or in the cloud. Many other LIMS, such as Exemplar (Laboratory Information Management System Software (LIMS) | Sapio Sciences 2020), MISO (2020), and NG6 (Mariette et al. 2012), also provide a variety of tools for the management of NGS experiments.

11.7 Other Issues to Consider

Apart from scientific, technical, and logistic challenges, there are other ancillary aspects that one must consider in managing NGS data. First, the acquisition, storage, and usage of personalized omics data may lead to unintentional legal, social, and ethical breaches; hence, appropriate framework should be put in place (Minkoff and Ecker 2008). For instance, it remains controversial as to whether first-degree relatives and insurance companies are eligible to access and use personal genomic information, in addition to possible genetic discrimination (Federal Register 2015). Other considerations include the consent process and the communication results and counseling (Fossey et al. 2018). The European Union introduced the General Data Protection Regulation (GDPR), introduced in 2016, to set specific provisions for the processing of sensitive data, such as technical and organizational measures on adapting pseudonymization which intended to deliver adequate protection rights and freedoms of data subjects (Sanchini and Marelli 2020).

Translation of big data into PM requires healthcare providers to possess knowledge in molecular life sciences, so that they are able to understand and interpret these big data reports, as well as to incorporate these data into disease treatment and prevention, and to deliver the appropriate knowledge to patients (Reference GH 2020). However, the current generation of clinicians and nursing staff often lack the background for managing EHRs, omics data, and biobanks, besides differentiating actionable genes and providing options for participants to receive limited results or to withdraw from the study at a later stage. It is therefore imperative to incorporate these cross-disciplinary components in tertiary education and continuing professional development. To facilitate knowledge sharing, it is also highly beneficial to standardize terminologies and nomenclatures, besides NGS data structures and formats in this emergent field. Besides, pioneering endeavors are often greeted with skepticism and resistance by the public due to the lack of exposure. As such, public education and campaigns are equally important for creating public awareness and acceptance.

Another compelling factor for health institutions is the cost. Storing and managing NGS datasets, EHRs, and other lab tests require hefty initial investment in high-performance computing, maintenance, and periodic upgrade, not to mention hefty investment in human capital and infrastructure upon scaling up to nationwide public health efforts. This will inevitably involve politics whereby public debates on health economics, crafting of new budgets, and policies will take place.

11.8 Conclusion and Future Perspectives

The improved reliability and lowering cost have rendered NGS being increasingly adopted in the clinics; and these have also further expanded the NGS “big data” (Pennell et al. 2019; Hulsen et al. 2019). Such advancement should have been

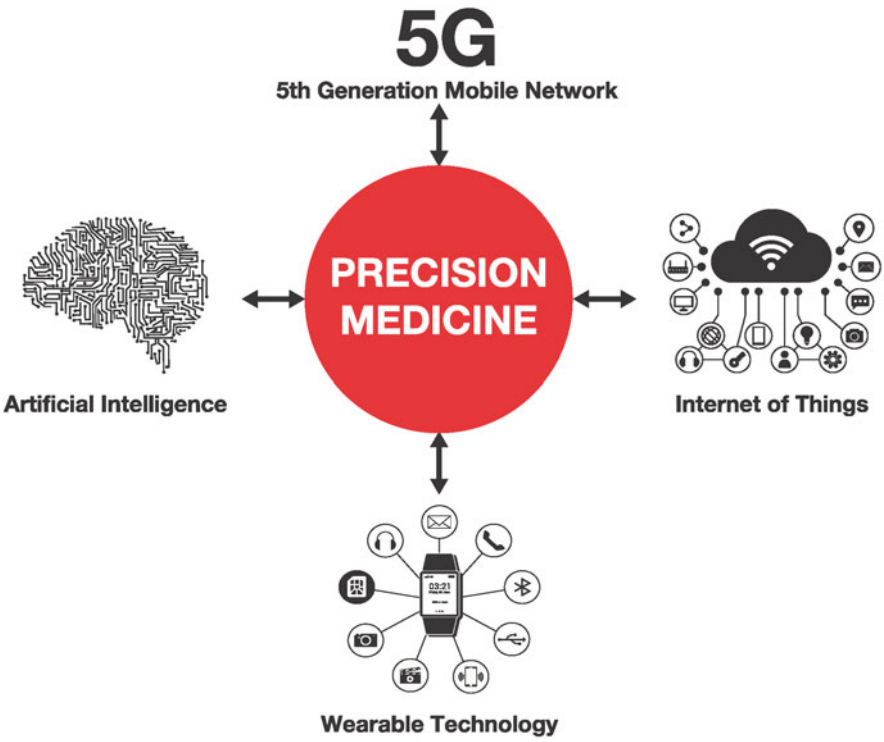

Fig. 11.4 Developments in the Fourth Industrial Revolution (4IR) that can improve the quality, quantity, variety, and connectivity of acquired NGS and phenotypic data for precision medicine

matched with an equivalent capacity in automated data analysis and interpretation, but NGS data are still mainly curated by bioinformaticians and subject matter experts manually. We envisage that a major revolution in precision medicine will be artificial intelligence (AI), an indispensable component of the Fourth Industrial Revolution (4IR) which will alleviate the bottleneck in NGS data analysis (Fig. 11.4). AI refers to a range of computational capabilities, i.e., rule-based computing, machine learning (ML), deep learning, natural language processing (NLP), and computer vision, that can facilitate and speed up tasks requiring human intelligence in reasoning, decision-making, as well as speech recognition, visual perception, or detection of data patterns.

Currently AI has begun to exhibit its prowess in NGS data analysis. Firstly, in the variant calling process whereby a sample nucleic acid sequence is compared to a reference sequence to detect any differences, manual annotations of variants often produce errors and biases (Hwang et al. 2015). AI algorithms such as DeepVariant can learn these biases from a single genome with a reference standard of variant calls and produce premium variant calls (Poplin et al. 2018). Meanwhile, in genome annotation and variant classification, prediction of both (i) the functional elements from primary DNA sequence and (ii) the effects of genetic variations on these

functional elements can also be facilitated by AI, as exemplified by algorithms such as combined annotation-dependent depletion (CADD) (Kircher et al. 2014), SpliceAI (Jaganathan et al. 2019), and DeepSEA (Bernstein et al. 2010). Thirdly, another process that can be expedited by AI is the mapping of phenotypes to genotypes. Every predicted pathogenic variant should preferably be matched to an expected disease phenotype to ease subsequent clinical diagnosis. Through an interdisciplinary area named computer vision, AI-based deep learning has been implemented on medical imaging data obtained from different diseases; and it has been reported to outperform expert pathologists and dermatologists (De Fauw et al. 2018; Bejnordi et al. 2017). A good example of such an AI-based facial image analysis algorithm is DeepGestalt (Gurovich et al. 2019). DeepGestalt can be incorporated into PEDIA, a genome interpretation system. PEDIA was able to analyze phenotypic features extracted from facial photographs to accurately prioritize candidate pathogenic variants for 105 different monogenic disorders across 679 individuals (Hsieh et al. 2019). Likewise, AI-based speech recognition has been applied to detect diseases such as chronic pharyngitis, which negatively affects speech ability (Li et al. 2019). Finally, NLP can be used to recognize and capture patterns in EHRs so that other patient-related data such as laboratory tests or family history can also be incorporated so as to predict diagnoses, to drive phenotype-informed genetic analysis, and to inform clinical decision-making (Clark et al. 2019).

Apart from AI, Internet of Things (IoT) is another upcoming trend in PM. At the moment, apart from mobile devices, various wearable smart devices such as smartwatches, fitness trackers, smart jewelry, smart clothing, and head-mounted displays are already in the consumer electronics market. These wearables are usually worn close to the skin surface so that the embedded sensors can continuously detect, analyze, and transmit signals of vital signs, ambient data, or physical activity to the users (Düking et al. 2018). Importantly, these constantly acquired data can be transmitted and uploaded to cloud storage in real time, so that they can provide phenotypic features that can be interpreted by AI. At the other end, this form of portable, real-time, on-site, and IoT-based analysis has also been realized in nanopore-based DNA/RNA sequencing, which has recently been commercialized (Liu et al. 2016). Together with the impending fifth-generation mobile network (5G) that is designed to connect virtually everyone and every device at an elevated speed and reliability, we therefore believe that healthcare will be completely transformed.

In short, current strategies for NGS data management in PM, as discussed in detail here, are still very much under development, and not many healthcare providers, even in the advanced countries, have even started to embrace NGS-based PM. Even then, the pace of technologies has moved so rapidly that NGS data strategies need to be redesigned from time to time in order to keep abreast of this age of 4IR. To remain future-compliant, new ideas such as AI, IoT, wearable technologies, and 5G communication will need to be assimilated into PM so that it can further benefit from the real-time phenotypic data that we have started to accumulate on a daily basis.

Acknowledgments Research in the authors' laboratories is supported by the Research University Grant (Geran Universiti Penyelidikan) of Universiti Kebangsaan Malaysia (No. GUP-2020-076 and GUP-2020-078). The authors would like to thank Ms. Shu Ning Low for the preparation of figures.

References

Acha P, Xandri M, Fuster-Tormo F, Palomo L, Xicoy B, Cabezón M, et al. Diagnostic and prognostic contribution of targeted NGS in patients with triple-negative myeloproliferative neoplasms. Am J Hematol. 2019;94:E264–9. Wiley-Liss Inc.; Available from: https://pubmed.ncbi.nlm.nih.gov/31321810/

Alonso CM, Llop M, Sargas C, Pedrola L, Panadero J, Hervás D, et al. Clinical utility of a next-generation sequencing panel for acute myeloid leukemia diagnostics. J Mol Diagn. 2019;21(2):228–40. Available from: https://pubmed.ncbi.nlm.nih.gov/30576870/

Anatskiy E, Ryan DP, Grüning BA, Arrigoni L, Manke T, Bönisch U. Parkour LiMs: High-quality sample preparation in next generation sequencing. Bioinformatics. 2019;35(8):1422–4. Available from: https://pubmed.ncbi.nlm.nih.gov/30239601/

Argento N. Institutional ELN/LIMS deployment. EMBO Rep. 2020;21(3) https://doi.org/10.15252/embr.201949862.

Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation. Nature. 2015;526:68–74. Available from: https://www.nature.com/articles/nature15393. Nature Publishing Group

Barillari C, Ottoz DSM, Fuentes-Serna JM, Ramakrishnan C, Rinn B, Rudolf F. OpenBIS ELN-LIMS: An open-source database for academic laboratories. Bioinformatics. 2016;32(4):638–40. Available from: https://academic.oup.com/bioinformatics/article/32/4/638/1743839

BaseSpace Clarity LIMS: Lab management and automation for genomics [Internet]. [cited 2020 Sep 14]. Available from: https://www.illumina.com/products/by-type/informatics-products/basespace-clarity-lims.html

Basik M, Aguilar-Mahecha A, Rousseau C, Diaz Z, Tejpar S, Spatz A, et al. Biopsies: next-generation biospecimens for tailoring therapy. Nat Rev Clin Oncol. 2013;10(8):437–50. https://doi.org/10.1038/nrclinonc.2013.101.

Bass AJ, Thorsson V, Shmulevich I, Reynolds SM, Miller M, Bernard B, et al. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014;513(7517):202–9. Available from: https://www.nature.com/articles/nature13480

Bejnordi BE, Veta M, Van Diest PJ, Van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. J Am Med Assoc. 2017;318(22):2199–210.

Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH roadmap epigenomics mapping consortium. Nat Biotechnol. 2010;28:1045–8.

Bianchi V, Ceol A, Ogier AGE, de Pretis S, Galeota E, Kishore K, et al. Integrated systems for NGS Data Management and Analysis: open issues and available solutions. Front Genet. 2016;7:75. https://doi.org/10.3389/fgene.2016.00075/abstract.

Blum A, Wang P, Zenklusen JC. SnapShot: TCGA-analyzed tumors. Cell. 2018;173(2):530. Cell Press; 2018 [cited 2020 Sep 13]. Available from: https://pubmed.ncbi.nlm.nih.gov/29625059/

Bylstra Y, Davila S, Lim WK, Wu R, Teo JX, Kam S, et al. Implementation of genomics in medical practice to deliver precision medicine for an Asian population. NPJ Genomic Med. 2019;4(1):12. https://doi.org/10.1038/s41525-019-0085-8.

Cała J, Xu Y, Wijaya EA, Missier P. From scripted HPC-based NGS pipelines to workflows on the cloud. In: Proceedings—14th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing, CCGrid. 2014. IEEE Computer Society; 2014. p. 694–700.

Cargill CF. Why standardization efforts fail. J Electron Publ. 2011;14(1) Available from: http://hdl.handle.net/2027/spo.3336451.0014.103

Chen Y, Lin Y, Yuan X, Shen B. LIMS and clinical data management. In: Advances in experimental medicine and biology. New York LLC: Springer; 2016. p. 225–39. Available from: https://pubmed.ncbi.nlm.nih.gov/27807749/.

Clark MM, Hildreth A, Batalov S, Ding Y, Chowdhury S, Watkins K, et al. Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation. Sci Transl Med. 2019;11(489)

Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med. 2015;372(9):793–5.

Conrads G, Abdelbary MMH. Challenges of next-generation sequencing targeting anaerobes. Anaerobe. 2019;58:47–52. Academic Press

Cowie MR, Blomster JI, Curtis LH, Duclaux S, Ford I, Fritz F, et al. Electronic health records to facilitate clinical research. Clin Res Cardiol. 2017;106(1):1–9.

Dander A, Pabinger S, Sperk M, Fischer M, Stocker G, Trajanoski Z. SeqBench: integrated solution for the management and analysis of exome sequencing data. BMC Res Notes. 2014;7(1):43. Available from: https://bmcresnotes.biomedcentral.com/articles/10.1186/1756-0500-7-43

De Block M. The hospital and its IT system: where it is right now and what it needs. In: Hospital logistics and e-Management. Wiley; 2019. p. 13–36. https://doi.org/10.1002/9781119670537.ch2.

De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med. 2018;24(9):1342–50.

Doestzada M, Vila AV, Zhernakova A, Koonen DPY, Weersma RK, Touw DJ, et al. Pharmacomicrobiomics: a novel route towards personalized medicine? Protein Cell. 2018;9(5):432–45.

Dubbink HJ, Atmodimedjo PN, Kros JM, French PJ, Sanson M, Idbaih A, et al. Molecular classification of anaplastic oligodendroglioma using next-generation sequencing: a report of the prospective randomized EORTC Brain Tumor Group 26951 phase III trial. Neuro Oncol. 2016;18(3):388–400. Available from: https://pubmed.ncbi.nlm.nih.gov/26354927/

Düking P, Achtzehn S, Holmberg HC, Sperlich B. Integrated framework of load monitoring by a combination of smartphone applications, wearables and point-of-care testing provides feedback that allows individual responsive adjustments to activities of daily living. Sensors (Switzerland). 2018;18(5)

Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57–74. Available from: http://encodeproject.org/ENCODE/

Durmaz AA, Karaca E, Demkow U, Toruner G, Schoumans J, Cogulu O. Evolution of genetic techniques: past, present, and beyond, vol. 2015. BioMed Research International. Hindawi Publishing Corporation; 2015.

Eckel-Passow JE, Lachance DH, Molinaro AM, Walsh KM, Decker PA, Sicotte H, et al. Glioma groups based on 1p/19q, *IDH*, and *TERT* promoter mutations in tumors. N Engl J Med. 2015;372(26):2499–508. https://doi.org/10.1056/NEJMoa1407279.

El-Kalioby M, Abouelhoda M, Krüger J, Giegerich R, Sczyrba A, Wall DP, et al. Personalized cloud-based bioinformatics services for research and education: use cases and the elastic HPC package. BMC Bioinformatics. 2012;13(Suppl. 17):S22. https://doi.org/10.1186/1471-2105-13-S17-S22.

ElRakaiby M, Dutilh BE, Rizkallah MR, Boleij A, Cole JN, Aziz RK. Pharmacomicrobiomics: the impact of human microbiome variations on systems pharmacology and personalized therapeutics. OMICS. 2014;18(7):402–14.

Endrullat C, Glökler J, Franke P, Frohme M. Standardization and quality management in next-generation sequencing. Appl Transl Genomics. 2016;10:2–9. Elsevier B.V.

Federal Register: 2015 Edition Health Information Technology (Health IT) Certification Criteria, 2015 Edition Base Electronic Health Record (EHR) Definition, and ONC Health IT Certification Program Modifications.

Fossey R, Kochan D, Winkler E, Pacyna JE, Olson J, Thibodeau S, et al. Ethical considerations related to return of results from genomic medicine projects: the eMERGE network (phase III) experience. J Pers Med. 2018;8(1)

Gao Q, Liang WW, Foltz SM, Mutharasu G, Jayasinghe RG, Cao S, et al. Driver fusions and their implications in the development and treatment of human cancers. Cell Rep. 2018;23(1):227–38. e3. Available from: https://pubmed.ncbi.nlm.nih.gov/29617662/

Gibbon GA. A brief history of LIMS. Lab Autom Inf Manag. 1996;32(1):1–5.

GNomEx. [cited 2020 Sep 14]. Available from: https://hci-bio-app.hci.utah.edu/gnomex/authenticate

Gonzalez-Garay ML. The road from next-generation sequencing to personalized medicine. Personal Med. 2014;11(5):523–44. Future Medicine Ltd.; 2014 [cited 2020 Sep 13]. p. 523–44. Available from: /pmc/articles/PMC4437232/?report=abstract

Grabiner BC, Nardi V, Birsoy K, Possemato R, Shen K, Sinha S, et al. A diverse array of cancer-associated MTOR mutations are hyperactivating and can predict rapamycin sensitivity. Cancer Discov. 2014;4(5):554–63. Available from: https://pubmed.ncbi.nlm.nih.gov/24631838/

Grimes SM, Ji HP. MendeLIMS: a web-based laboratory information management system for clinical genome sequencing. BMC Bioinformatics. 2014;15(1):290. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-15-290

Gullapalli RR. Evaluation of commercial next-generation sequencing bioinformatics software solutions. J Mol Diagn. 2020;22(2):147–58.

Guo P, Peterson R, Paukstelis P, Wang J. Cloud-based life sciences manufacturing system: integrated experiment management and data analysis via Amazon web services. Cham: Springer; 2020. p. 149–59. https://doi.org/10.1007/978-3-030-30967-1_14.

Gurovich Y, Hanani Y, Bar O, Nadav G, Fleischer N, Gelbman D, et al. Identifying facial phenotypes of genetic disorders using deep learning. Nat Med. 2019;25(1):60–4.

Gutowska-Ding MW, Deans ZC, Roos C, Matilainen J, Khawaja F, Brügger K, et al. One byte at a time: evidencing the quality of clinical service next-generation sequencing for germline and somatic variants. Eur J Hum Genet. 2020;28(2):202–12. https://doi.org/10.1038/s41431-019-0515-1.

Hart EM, Barmby P, LeBauer D, Michonneau F, Mount S, Mulrooney P, et al. Ten simple rules for digital data storage. PLOS Comput Biol. 2016;12(10):e1005097.

Hou T, Wang T, Shen D, Lu Z, Liu Y. Autonomous security mechanisms for high-performance computing systems: review and analysis. In: Adaptive autonomous secure cyber systems; 2020.

Hsieh TC, Mensah MA, Pantel JT, Aguilar D, Bar O, Bayat A, et al. PEDIA: prioritization of exome data by image analysis. Genet Med. 2019;21(12):2807–14.

Hulsen T, Jamuar SS, Moody AR, Karnes JH, Varga O, Hedensted S, et al. From big data to precision medicine. Front Med. 2019;6:1–14.

Hwang S, Kim E, Lee I, Marcotte EM. Systematic comparison of variant calling pipelines using gold standard personal exome variants. Sci Rep. 2015;5(1):1–8.

Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting splicing from primary sequence with deep learning. Cell. 2019;176(3):535–48. e24

Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet. 2012;13(6):395–405.

Junaid S, Jangda Z. Successful deployment of a laboratory information management system LIMs; striding towards modern, paperless labs. In: International Petroleum Technology Conference 2020, IPTC 2020. International Petroleum Technology Conference (IPTC); 2020.

Kho AN, Rasmussen LV, Connolly JJ, Peissig PL, Starren J, Hakonarson H, et al. Practical challenges in integrating genomic data into the electronic health record. Genet Med. 2013;15(10):772–8.

Kircher M, Witten DM, Jain P, O'roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46(3):310–5.

Kohane IS. Using electronic health records to drive discovery in disease genomics. Nat Rev Genet. 2011;12(6):417–28. https://doi.org/10.1038/nrg2999.

Laboratory Information Management System Software (LIMS) | Sapio Sciences. [cited 2020 Sep 14]; Available from: https://www.sapiosciences.com/lab-management-software

Langhof H, Kahrass H, Illig T, Jahns R, Strech D. Current practices for access, compensation, and prioritization in biobanks. Results from an interview study. Eur J Hum Genet. 2018;26(11): 1572–81. Available from: https://pubmed.ncbi.nlm.nih.gov/30089824

Lemmon P, Vance, Jia Y, Shi Y, Douglas Holbrook S, Bixby JL, Buchser W. Challenges in small screening laboratories: implementing an on-demand laboratory information management system. Comb Chem High Throughput Screen. 2011;14(9):742–8.

Li Z, Huang J, Hu Z. Screening and diagnosis of chronic pharyngitis based on deep learning. Int J Environ Res Public Health. 2019;16(10)

Liu Z, Wang Y, Deng T, Chen Q. Solid-state nanopore-based DNA sequencing technology. J Nanomater. 2016;2016. Hindawi Limited

Malsagova K, Kopylov A, Stepanov A, Butkova T, Sinitsyna A, Izotov A, et al. Biobanks: a platform for scientific and biomedical research. Diagnostics (Basel, Switzerland). 2020;10(7)

Mangul S, Mosqueiro T, Abdill R, Duong D, Mitchell K, Sarwal V, et al. Challenges and recommendations to improve installability and archival stability of omics computational tools. bioRxiv. 2018:452532. https://doi.org/10.1101/452532.

Manrai AK, Kohane IS. Chapter 11: Bioinformatics and precision medicine. In: Sheikh A, Cresswell KM, Wright A, Bates DWBT-KA, editors. Key advances in clinical informatics. Academic Press; 2017. p. 145–60. Available from: http://www.sciencedirect.com/science/article/pii/B978012809523200011X.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437(7057):376–80. Available from: https://pubmed.ncbi.nlm.nih.gov/16056220

Mariette J, Escudié F, Allias N, Salin G, Noirot C, Thomas S, et al. NG6: Integrated next generation sequencing storage and processing environment. BMC Genomics. 2012;13(1):462. Available from: http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-13-462

Maróti Z, Boldogkői Z, Tombácz D, Snyder M, Kalmár T. Evaluation of whole exome sequencing as an alternative to BeadChip and whole genome sequencing in human population genetic analysis. BMC Genomics. 2018;19(1):778. https://doi.org/10.1186/s12864-018-5168-x.

Martinez-Martin N, Magnus D. Privacy and ethical challenges in next-generation sequencing. Expert Rev Precis Med Drug Dev. 2019;4(2):95–104.

Matthijs G, Souche E, Alders M, Corveleyn A, Eck S, Feenstra I, et al. Guidelines for diagnostic next-generation sequencing. Eur J Hum Genet. 2016;24:2–5. Available from: https://www.nature.com/articles/ejhg2015226; Nature Publishing Group

McLellan AS, Dubin RA, Jing Q, Broin PÓ, Moskowitz D, Suzuki M, et al. The Wasp System: an open source environment for managing and analyzing genomic data. Genomics. 2012;100(6): 345–51.

Mehandziska S, Stajkovska A, Stavrevska M, Jakovleva K, Janevska M, Rosalia R, et al. Workflow for the implementation of precision genomics in healthcare. Front Genet. 2020;11:619.

Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc. 2010;2010(6):pdb.prot5448.

Minkoff H, Ecker J. Genetic testing and breach of patient confidentiality: law, ethics, and pragmatics. Am J Obstet Gynecol. 2008;198(5):498.e1–498. e4

MISO | Earlham Institute [cited 2020 Sep 14]. Available from: https://www.earlham.ac.uk/miso/

Moore HM, Compton CC, Alper J, Vaught JB. International approaches to advancing biospecimen science. Cancer Epidemiol Biomarkers Prev. 2011;20(5):729–32. Available from: http://cebp.aacrjournals.org/content/20/5/729.abstract

Morash M, Mitchell H, Beltran H, Elemento O, Pathak J. The role of next-generation sequencing in precision medicine: a review of outcomes in oncology. J Pers Med. 2018;8(3):30.

Na K, Kim HS, Shim HS, Chang JH, Kang SG, Kim SH. Targeted next-generation sequencing panel (TruSight Tumor 170) in diffuse glioma: a single institutional experience of 135 cases. J Neurooncol. 2019;142(3):445–54. Available from: https://pubmed.ncbi.nlm.nih.gov/30710203/

Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Informatics Assoc. 2011;18(5):544–51.
Nagai A, Hirata M, Kamatani Y, Muto K, Matsuda K, Kiyohara Y, et al. Overview of the BioBank Japan Project: study design and profile. J Epidemiol. 2017;27(3S):S2–8. Available from: https://pubmed.ncbi.nlm.nih.gov/28189464
Navale V, Bourne PE. Cloud computing applications for biomedical science: a perspective. PLOS Comput Biol. 2018;14(6):e1006144. https://doi.org/10.1371/journal.pcbi.1006144.
NCI Best Practices for Biospecimen Resources. 2016.
Oberg JA, Glade Bender JL, Sulis ML, Pendrick D, Sireci AN, Hsiao SJ, et al. Implementation of next generation sequencing into pediatric hematology-oncology practice: moving beyond actionable alterations. Genome Med. 2016;8(1) Available from: https://pubmed.ncbi.nlm.nih.gov/28007021/
Ocaña K, De Oliveira D. Parallel computing in genomic research: advances and applications. In: Advances and applications in bioinformatics and chemistry, vol. 8. Dove Medical Press Ltd; 2015. p. 23–35. [cited 2020 Sep 1]. Available from: /pmc/articles/PMC4655901/?report=abstract.
Ollier W, Sprosen T, Peakman T. UK Biobank: from concept to reality. Pharmacogenomics. 2005;6(6):639–46.
Patel RK, Jain M, NGS QC. Toolkit: a toolkit for quality control of next generation sequencing data. PLoS One. 2012;7(2):e30619. https://doi.org/10.1371/journal.pone.0030619.
Paul S, Gade A, Mallipeddi S. The state of cloud-based biospecimen and biobank data management tools. Biopreserv Biobank. 2017;15(2):169–72. https://doi.org/10.1089/bio.2017.0019.
Pennell NA, Mutebi A, Zhou Z-Y, Ricculli ML, Tang W, Wang H, et al. Economic impact of next-generation sequencing versus single-gene testing to detect genomic alterations in metastatic non–small-cell lung cancer using a decision analytic model. JCO Precis Oncol. 2019;3:1–9.
Pennisi E. Genomics. Semiconductors inspire new sequencing technologies, vol. 327. Science (New York, N.Y.); 2010. p. 1190.
Peplow M. The 100,000 genomes project. BMJ. 2016;353:i1757. Available from: https://www.bmj.com/content/353/bmj.i1757
Pérez-Wohlfeil E, Torreno O, Bellis LJ, Fernandes PL, Leskosek B, Trelles O. Training bioinformaticians in high performance computing. Heliyon. 2018;4(12):1057. Available from: /pmc/articles/PMC6299036/?report=abstract
Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, et al. A universal snp and small-indel variant caller using deep neural networks. Nat Biotechnol. 2018;36(10):983.
Porreca GJ. Genome sequencing on nanoballs. Nat Biotechnol. 2010;28:43–4.
Proctor LM, Creasy HH, Fettweis JM, Lloyd-Price J, Mahurkar A, Zhou W, et al. The integrative human microbiome project. Nature. 2019;569(7758):641–8. Available from: https://pubmed.ncbi.nlm.nih.gov/31142853/
Rafid Raeen M. How laboratory informatics has impacted healthcare overall. Appl Res Proj. 2018; Available from: https://dc.uthsc.edu/hiimappliedresearch/54
Reference GH. Help Me Understand Genetics Precision Medicine [Internet]. 2020. Available from: https://ghr.nlm.nih.gov/%5Cn
Rhoads A, Au KF. PacBio sequencing and its applications. Genomics Proteomics Bioinformatics. 2015;13(5):278–89. Available from: http://www.sciencedirect.com/science/article/pii/S1672022915001345
Roberts S, Julius M. Precision medicine: now, not when. Healthc Manag Forum. 2016;29(4):158–61. Available from: https://pubmed.ncbi.nlm.nih.gov/27278139/. [cited 2020 Sep 13]
Robles-Espinoza CD, Adams DJ. Cross-species analysis of mouse and human cancer genomes. Cold Spring Harb Protoc. 2014, 2014;(4):350–8. Available from: http://cshprotocols.cshlp.org/content/2014/4/pdb.top078824.full
Roy S, Coldren C, Karunamurthy A, Kip NS, Klee EW, Lincoln SE, 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. Elsevier B.V.

Sanchini V, Marelli L. Data protection and ethical issues in European P5 eHealth. P5 eHealth: An Agenda Heal Technol Futur. 2020:173–89.

Sankar PL, Parker LS. The Precision Medicine Initiative's All of Us Research Program: An agenda for research on its ethical, legal, and social issues. Genet Med. 2017;19(7):743–50. Nature Publishing Group; Available from: https://www.nih.gov/research-training/

Saunders G, Baudis M, Becker R, Beltran S, Béroud C, Birney E, et al. Leveraging European infrastructures to access 1 million human genomes by 2022. Nat Rev Genet. 2019;20(11): 693–701. Available from: www.nature.com/nrg

Scheuner MT, De Vries H, Kim B, Meili RC, Olmstead SH, Teleki S. Are electronic health records ready for genomic medicine? Genet Med. 2009;11(7):510–7.

Scholtalbers J, Rößler J, Sorn P, de Graaf J, Boisguérin V, Castle J, et al. Galaxy LIMS for next-generation sequencing. Bioinformatics. 2013;29(9):1233–4. Available from: http://bcbio.wordpress.com/2011/01/11/next

Setia N, Agoston AT, Han HS, Mullen JT, Duda DG, Clark JW, et al. A protein and mRNA expression-based classification of gastric cancer. Mod Pathol. 2016;29(7):772–84. Available from: https://pubmed.ncbi.nlm.nih.gov/27032689/

Sinha R, Abnet CC, White O, Knight R, Huttenhower C. The microbiome quality control project: Baseline study design and future directions. Genome Biol. 2015;16(1) Available from: https://pubmed.ncbi.nlm.nih.gov/26653756/

Smith DR. Buying in to bioinformatics: An introduction to commercial sequence analysis software. Brief Bioinform. 2014;16(4):700–9. Available from: /pmc/articles/PMC4501248/?report=abstract

Stoll K, Kubendran S, Cohen SA. The past, present and future of service delivery in genetic counseling: keeping up in the era of precision medicine. Am J Med Genet Part C Semin Med Genet. 2018;178(1)24–37. doi: https://doi.org/10.1002/ajmg.c.31602

Stumptner C, Sargsyan K, Kungl P, Zatloukal K. Crucial role of high quality biosamples in biomarker development. In: Carini C, Fidock M, van Gool A, editors. Handbook of biomarkers and precision medicine. CRC Press Taylor & Francis Group; 2019. p. 128–34.

Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2504 human genomes. Nature. 2015;526(7571):75–81. Available from: https://www.nature.com/articles/nature15394

Tagger B. An Introduction and Guide to Successfully Implementing a LIMS (Laboratory Information Management System), 2011.

The Integrative HMP (iHMP) Research Network Consortium. The Integrative Human Microbiome Project: Dynamic analysis of microbiome-host omics profiles during periods of human health and disease corresponding author. Cell Host Microbe. 2014;16:276–89. Available from: https://pubmed.ncbi.nlm.nih.gov/25211071/. Cell Press

The Integrative Human Microbiome Project. Nature. 2019;569(7758):641–8.

Turnbull C, Scott RH, Thomas E, Jones L, Murugaesu N, Pretty FB, et al. The 100,000 Genomes Project: bringing whole genome sequencing to the NHS. BMJ. 2018:361. Available from: https://pubmed.ncbi.nlm.nih.gov/29691228/

Tyson JR, O'Neil NJ, Jain M, Olsen HE, Hieter P, Snutch TP. MinION-based long-read sequencing and assembly extends the Caenorhabditis elegans reference genome. Genome Res. 2017;28(2): 266–74. Available from: https://pubmed.ncbi.nlm.nih.gov/29273626

Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, et al. A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Res. 2008;18(7):1051–63. Available from: https://pubmed.ncbi.nlm.nih.gov/18477713

van Noord PAH. Banking of urine sediments as DNA source in epidemiologic studies. Epidemiology. 2003;14(2) Available from: https://journals.lww.com/epidem/Fulltext/2003/03000/Banking_of_Urine_Sediments_as_DNA_Source_in.23.aspx

Van Rossum T, Tripp B, Daley D. SLIMS-a user-friendly sample operations and inventory management system for genotyping labs. Bioinformatics. 2010;26(14):1808–10. Available from: https://academic.oup.com/bioinformatics/article/26/14/1808/178330

Venco F, Vaskin Y, Ceol A, Muller H. SMITH: A LIMS for handling next-generation sequencing workflows. BMC Bioinformatics. 2014;15(Suppl. 14):S3. Available from: http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-15-S14-S3

Volckmar AL, Leichsenring J, Kirchner M, Christopoulos P, Neumann O, Budczies J, et al. Combined targeted DNA and RNA sequencing of advanced NSCLC in routine molecular diagnostics: Analysis of the first 3000 Heidelberg cases. Int J Cancer. 2019;145(3):649–61. Available from: https://pubmed.ncbi.nlm.nih.gov/30653256/

Wagle N, Grabiner BC, Van Allen EM, Hodis E, Jacobus S, Supko JG, 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(5):546–53. Available from: https://pubmed.ncbi.nlm.nih.gov/24625776/

Wandelt S, Rheinländer A, Bux M, Thalheim L, Haldemann B, Leser U. Data management challenges in next generation sequencing. Datenbank-Spektrum 2012;12(3):161–171. Available from: https://link.springer.com/article/10.1007/s13222-012-0098-2

Woo JS, Lu DY. Procurement, transportation, and storage of saliva, buccal swab, and oral wash specimens. Methods Mol Biol. 2019;1897:99–105.

Wu PY, Cheng CW, Kaddi CD, Venugopalan J, Hoffman R, Wang MD. Omic and Electronic health record big data analytics for precision medicine. IEEE Trans Biomed Eng. 2017;64(2):263–73. Available from: https://pubmed.ncbi.nlm.nih.gov/27740470/

Zacher A, Kaulich K, Stepanow S, Wolter M, Köhrer K, Felsberg J, et al. Molecular diagnostics of gliomas using next generation sequencing of a glioma-tailored gene panel. Brain Pathol. 2017;27(2):146–59. Available from: https://pubmed.ncbi.nlm.nih.gov/26919320/

Zhang H, He L, Cai L. Transcriptome sequencing: RNA-Seq. Methods Mol Biol. 2018;1754:15–27.