OXFORD

Briefings in Bioinformatics, 20(3), 2019, 752–766

doi: 10.1093/bib/bbx144
Advance Access Publication Date: 25 October 2017
Paper

Precision medicine needs pioneering clinical bioinformaticians

Gonzalo Gómez-López, Joaquín Dopazo, Juan C. Cigudosa, Alfonso Valencia and Fátima Al-Shahrour

Corresponding author. Fátima Al-Shahrour, Bioinformatics Unit, Spanish National Cancer Research Centre (CNIO), C/Melchor Fernández Almagro, 3. ES-28029 Madrid, Spain. Tel.: +34 91 732 80 00; E-mail: falshahrour@cnio.es

Abstract

Success in precision medicine depends on accessing high-quality genetic and molecular data from large, well-annotated patient cohorts that couple biological samples to comprehensive clinical data, which in conjunction can lead to effective therapies. From such a scenario emerges the need for a new professional profile, an expert bioinformatician with training in clinical areas who can make sense of multi-omics data to improve therapeutic interventions in patients, and the design of optimized basket trials. In this review, we first describe the main policies and international initiatives that focus on precision medicine. Secondly, we review the currently ongoing clinical trials in precision medicine, introducing the concept of 'precision bioinformatics', and we describe current pioneering bioinformatics efforts aimed at implementing tools and computational infrastructures for precision medicine in health institutions around the world. Thirdly, we discuss the challenges related to the clinical training of bioinformaticians, and the urgent need for computational specialists capable of assimilating medical terminologies and protocols to address real clinical questions. We also propose some skills required to carry out common tasks in clinical bioinformatics and some tips for emergent groups. Finally, we explore the future perspectives and the challenges faced by precision medicine bioinformatics.

Key words: precision medicine; computing infrastructures; clinical bioinformatics; training; clinical bioinformatician; genomic report

Precision medicine in the real world: the dress rehearsals

The paradigm of precision medicine is defined by combining the use of population-based molecular profiling, clinical data, epidemiological information and other types of data to make clinical decisions that are tailored to individual patients [1]. The potential advantages of this approach, both for patients and doctors, include more accurate diagnosis and treatments, safer drug prescription, better disease prevention and consequently, a reduction in healthcare costs. The integration of genomics into routine clinical practice requires systems and workforces that are equipped and prepared to handle the scale and complexity of genomic data. As such, bioinformatics plays an essential role in

Gonzalo Gómez-López is a senior computational biologist at the Spanish National Cancer Research Centre, and assistant professor of translational genomics and bioinformatics at the School of Medicine of the Autónoma University (Madrid). He is interested in cancer genome interpretation, transcriptomics and immunotherapy.
Joaquín Dopazo is the director of the Clinical Bioinformatics Area of the Fundación Progreso y Salud (Seville). His research focuses on translational genomic data analysis for precision medicine, with an emphasis on systems medicine applications.
Juan Cruz Cigudosa is a clinical geneticist with experience in genomics for research and clinical applications. He is the current president of the Spanish Human Genetics Association and scientific director of NIMGenetics.
Alfonso Valencia is the director of the Spanish National Bioinformatics Institute (INB) and the Life Sciences department at the Barcelona Supercomputing Centre. His main research goals are to analyse the function and interactions of cancer-related proteins, and to develop novel computational methods to examine, represent and interpret cancer genome information.
Fátima Al-Shahrour is head of the Bioinformatics Unit at Spanish National Cancer Research Centre. She is applying computational methods to precision medicine to interpret cancer genomes, for drug repositioning and for the prediction of anticancer therapies.
Submitted: 17 July 2017; **Received (in revised form):** 14 September 2017

providing the elements required for the processing, visualization and interpretation of a patient's multi-omics profiles, and for the integration of these profiles with clinical data to gain a mechanistic understanding of their disease, thereby facilitating more personalized treatment [2]. Common bioinformatics tasks in a precision medicine scenario include the implementation and execution of well-established and reproducible workflows to process a patient's omics data, applying computational methods to detect altered genes (mutated, amplified/deleted, altered expression, etc.), to interpret the biological and clinical impact of such alterations, to establish therapeutic guidance based on the patient's genomic profile and for health record data mining to achieve knowledge-driven clinical assessment (Figure 1). In theory, all these tasks enable genome-based reports to be generated that can eventually stratify patients to facilitate clinical decision-making. The aforementioned bioinformatics tasks must be supported by robust and stable technological platforms, and computational infrastructures for data storage, data privacy and protection, which also contemplate protocols for server maintenance and pipeline management [3].

The focus on human health and disease brings new challenges and requirements to bioinformaticians, particularly given the volume, complexity, heterogeneity and nature of the data. Computational systems biomedicine is an emerging discipline [4] that aims to provide the computing methodologies, communication technologies and tools to tackle the problems derived from the complex nature of many human health issues and diseases. Although examples can be found in basic research areas, the application of systems medicine to the clinic is still relatively limited. Importantly, bioinformaticians face becoming novel stakeholders in the healthcare sector who will collaborate closely with physicians in clinical decision-making, henceforth becoming clinical bioinformaticians.

Translating cancer genomes into the clinic

Although the model of precision medicine may apply to many diseases, cancer is the disease for which it is clearly most advanced at this time. Precision oncology has incorporated the study of human cancers by genome sequencing and/or other genome-based technologies, proving that many tumours harbour hundreds of gene mutations and/or copy number changes [5]. In this scenario, the genetic heterogeneity arising from the bioinformatics analysis of tumour genome profiles indicates that the majority of cancers are not single diseases but rather, they are an array of disorders with distinct molecular mechanisms and where there is clinical variation between individuals [6–9]. Such genetic variation in tumours includes essential information to guide the diagnosis and treatment of cancer patients [10]. However, most clinical trials currently evaluate the efficacy and safety of a new drug by analysing its effects on largely unselected populations of patients, ignoring their genomic profile, a characteristic that might indeed help to identify the patients that are most likely to respond to the treatment [11]. Consequently, precision oncology has been incorporated into clinical trials. Moreover, molecular profiling of cancer has bolstered the concept of 'basket trials', a new and evolving type of clinical trial designed around the hypothesis that the presence of a molecular marker predicts the response to a targeted therapy independently of the tumour type [12].

Large-scale cancer genome consortia have emerged for nearly all major cancer types to comprehensively characterize the genomes of thousands of cases, including The Cancer Genome Atlas (TCGA; https://cancergenome.nih.gov/) and The International Cancer Genome Consortium (ICGC; http://icgc.org/) [13]. Unfortunately, these large-scale genomic projects were launched without complete and standardized clinical information of the donors, such as the treatments received or the patient's medical history, making it difficult to identify new preventative and predictive genomic biomarkers that would aid the design of biomarker-driven clinical trials. Therefore, an important observation drawn from the completion of these projects is the importance of data sharing as an essential way to link genomic data with high-quality and standardized clinical information when attempting to identify genotype–phenotype associations [14, 15]. According to the 2016 Precision Medicine Essential Brief, >60% of respondents indicated that the most significant challenges in precision medicine are gaining access to clinical data, the integration of clinical data systems and the integration of clinical and genomic data [16]. Indeed, the best way to achieve clinical and genomic data sharing is currently a subject of intense debate [17–19].

Precision medicine national initiatives

In recent years, many countries have implemented different precision medicine initiatives (PMIs) at a national level, including the USA, China, Australia, Qatar, South Korea and in Europe,

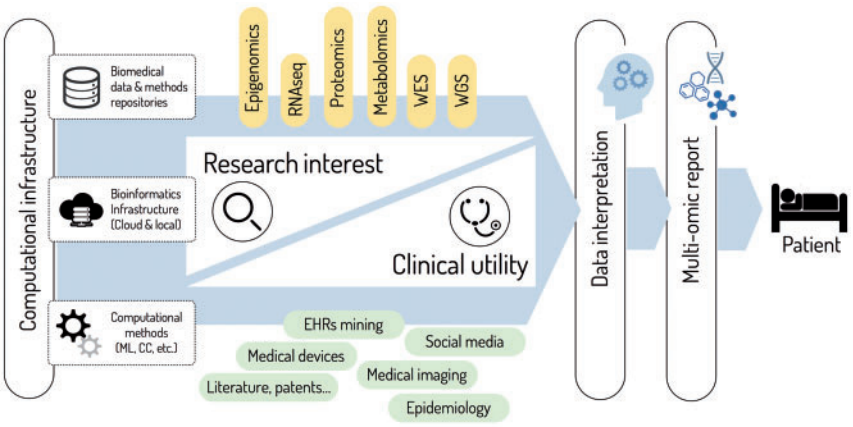

Figure 1. Precision medicine workflow: from data to patient care. Precision medicine requires computational infrastructures to efficiently store and process data on patient genotypes and phenotypes. The biological and clinical interpretation of such data is converted into an integral multi-omics report that will support clinical decision-making: ML, Machine Learning; CC, Cognitive Computing; EHRs, Electronic Health Records.

England, France, Finland, Denmark, the Netherlands and Germany (Table 1). The US PMI, known as 'All of Us', aims to initiate a paradigm shift for modern medicine by increasing population-based genome sequencing, and linking it with clinical data to understand and determine how best to prevent or treat disease. Notably, 'All of Us' will encourage open data sharing, allowing access to patients and researchers [20]. Another relevant personalized medicine approach is the '100 000 Genomes Project' in the UK, which is generating significant amounts of genomic data from rare diseases and cancer to inform clinical decision-making. Here, it is important to highlight how governing bodies, including research and health policy makers, are working together to support these initiatives, establishing international consortiums that provide policy recommendations, as well as research and training activities designed to exploit the potential of personalized medicine to the full. Some examples include the International Consortium for Personalized Medicine (ICPerMed; http://www.icpermed.eu/) or European Alliance for Personalized Medicine (EAPM; http://euapm.eu).

Both national and international PMIs can be found at different stages of development. Some interesting efforts are currently being made to evaluate the social and economic benefit of personalized medicine in the healthcare sector. However, for all of these initiatives, there are significant challenges that need to be addressed to aid the broader implementation of precision medicine in the hospital setting: organizational, ethical and regulatory challenges; shortfalls in the generation of evidence; challenges in data sharing and infrastructure; the slow uptake of genomic information into clinical care and research; the economics of precision medicine [21]. Legislation to protect genomic data [22], the development of decision-making tools, preparing work groups and achieving greater patient/clinician engagement and trust [23] are all issues that must be addressed before implementing solutions to these challenges. Moreover, multidisciplinary efforts will be necessary to overcome the future challenges to the implementation of precision medicine, where bioinformatics will certainly play a key role.

International consortiums for precision medicine: unity is strength

Data mining and the potential for data sharing are key aspects of a number of recent high-profile genomics/personalized medicine initiatives with significant clinical potential (Table 1). For instance, the International Cancer Genome Consortium for Medicine (ICGCmed; http://icgcmed.org/) aims to address the shortcomings in standardization and in quality control of the clinical information available for patients enrolled by ICGC or TCGA. ICGCmed will collect a much richer genomics data set associated with clinical and health-related information, and that regarding response to therapies, making it more valuable for personalized medicine. Similarly, the Genomics Evidence Neoplasia Information Exchange (GENIE) project [24], supported by the American Association for Cancer Research, recently released almost 19 000 de-identified genomic records along with some clinical data collected from cancer patients as part of routine patient care. Although progress has been made, the successful execution of these projects will rely on effective global strategies for sharing disease-related clinical data and on these data complying with standards, which in general are not yet fully defined.

A large number of international initiatives have emerged to develop guidelines for the responsible collection, curation, sharing and use of patient's clinical and genomic data, and to create a harmonious approach to data sharing between the existing databases. The Global Alliance for Genomics and Health (GA4GH; https://genomicsandhealth.org/) [25] is dedicated to create interoperable technical standards to manage and share genomic and clinical data. Interestingly, the GA4GH website provides a full catalogue of worldwide genomic data initiatives, data-sharing efforts, databases and repositories, international genomics research consortia and projects and other genomics data resources [26]. The GA4GH roadmap includes the catalysis of data sharing projects, and resolving particular needs in PMIs by applying working products and demonstration projects. Some successful examples include the BRCA challenge project (http://brcaexchange.org/), which provides an efficient platform with clinical information of BRCA mutations collected from patients and their phenotypic characteristics, and Matchmaker Exchange (http://www.matchmakerexchange.org/), a federated network of databases whose goal is to find genetic causes of rare diseases by matching similar phenotypic and genotypic profiles [27]. For its medical relevance, the ClinVar Initiative that is hosted at the NIH site is also being accepted as a common shared repository that links phenotypes and genomic variation with supporting evidence [28].

Precision medicine requires advanced technologies and processes to collect, manage and analyse data, and to provide rapid and precise decision support within a clinical and public health context. Large computing resources already exist, yet they need to be fully public and operate according to global standards. ELIXIR is Europe's leading life science infrastructure responsible for managing access to the massive amounts of data generated every day by publicly funded research (https://www.elixir-europe.org/). ELIXIR is building the technical infrastructure required by researchers to discover, combine and exchange human data via controlled access and in the context of the European Open Science Cloud initiative, while complying with data privacy and data security requirements. The backbone of the Human Data Use Case is the European Genome-phenome Archive (https://ega.crg.eu/) and the associated facilities, which include the GA4GH Beacon Project (https://beacon-network.org/). These initiatives aim to develop an open sharing platform as a simple public web service to aid genomic data centres make their data more 'discoverable', without revealing any sensitive information. As part of the global trend towards data sharing and open access to data repositories, the National Cancer Institute's Genomic Data Commons constitutes a valuable research resource. This initiative provides significant quantities of data from NCI sponsored research, with over four petabytes of data initially being released to the global cancer community to allow comprehensive mining of this information [29].

Testing the paradigm: precision clinical trials

The introduction of precision medicine at academic medical centres and in multi-hospital healthcare systems is ongoing. A recent survey conducted by the Healthcare Information and Management Systems Society (http://www.himssanalytics.org/) focused on larger hospitals and healthcare systems in the USA, and only 29% of the 137 respondents declared that they were engaged in some kind of precision medicine research. It is expected that current policies and initiatives driven by the 'All of Us' research programme in the USA and by the ICPerMed consortium in EU will facilitate and expand the implementation of precision medicine in Western countries. Until they do, some current cutting-edge projects are already setting the stage for the application of precision medicine via clinical trials.

Table 1. National and international PMIs and consortiums

Initiative	Description	Funding/Partners	URL
PMI's cohort program—'All of us' Research Program	PMI was launched in 2015 to make advances in tailoring medical care to the individual. The program will collect genetic and health data from one million people.	NIH	https://allofus.nih.gov/
MyCode Community Health Initiative	As part of PMI, Geisinger's MyCode Community Health Initiative represents the largest study in the United States with EHRs linked to large-scale DNA sequencing data. Nearly 150 000 patient participants have already been signed up.	NIH/Geisinger Health System hospitals.	https://www.geisinger.edu/research/departments-and-centers/genomic-medicine-institute/mycode-health-initiative
ICGCmed	ICGCmed will link the wealth of genomic data already amassed across the cancer spectrum, with new genomic data being generated, and with clinical and health information that includes lifestyle, patient history, cancer diagnostic data and response to and survival following therapy.	Funding agencies in Asia, Australia, Europe, North America and South America, supporting 88 projects in 17 jurisdictions (16 countries and the European Union), to study over 25 000 tumour genomes in 26 different tumour types.	https://icgcmed.org/
GenomeAsia100k	A non-profit consortium collaborating to sequence and analyse 100 000 Asian individuals	Macrogen (South Korea) and MedGenome.	http://www.genomeasia100k.com/
Project GENIE	GENIE is a multi-phase, multi-year, international data-sharing project that catalyses precision oncology through the development of a regulatory-grade registry that aggregates and links clinical-grade cancer genomic data with clinical outcomes from tens of thousands of cancer patients treated at multiple international institutions.	AACR/Dana-Farber Cancer Institute (USA), Gustave Roussy Cancer Campus (France), NKI (The Netherlands), Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins (USA), Memorial Sloan Kettering Cancer Center (USA), Princess Margaret Cancer Centre (Canada), University of Texas MD Anderson Cancer Center (USA), Vanderbilt-Ingram Cancer Center (USA)	http://www.aacr.org/Research/Research/Pages/aacr-project-genie.aspx
Worldwide Innovative Networking (WIN) Consortium in personalized cancer medicine	WIN was created to accelerate the pace and reduce the cost of translating novel cancer treatments to the bedside through worldwide clinical trials and research projects	Global initiative, headquartered in Paris, includes 35 institutional members.	http://winconsortium.org/
EAPM	The EAPM initiative was created to bring together European healthcare experts and patient advocates involved with major chronic diseases.	EU	http://euapm.eu/
100, 000 Genomes Project	This project will sequence 100 000 genomes from around 70 000 patients, combining genomic sequence data with medical records. Participants are NHS patients with a rare disease, plus their families, and patients with cancer.	NHS Genomics England, UK	https://www.genomicsengland.co.uk/
The ICPerMed	ICPerMed brings together over 30 European and international partners representing ministries, funding agencies and the European Commission. ICPerMed provides a platform to initiate and support communication and exchange on personalized medicine research, funding and implementation.	European Commission/EU countries and others	http://www.icpermed.eu/index.php
France Médicine Génomique 2025	Among the objectives of France Médicine Génomique is to perform approximately 10 000 WGS corresponding to 20 000 patients with rare diseases and their families, and 50 000 patients with metastatic or refractory cancers.	French Government	http://presse.inserm.fr/wp-content/uploads/2016/06/Plan-France-me%CC%81decine-ge%CC%81nomique-2025.pdf

Continued

Table 1. (continued)

Initiative	Description	Funding/Partners	URL
Estonian Genome Project (EGP)	The EGP is a large population-based databank that was established with health records and biological samples from a large portion of the population. Its aim is for these data to be used in biomedical and genetic research to improve future public healthcare in Estonia.	Estonian Government	http://www.geenivaramu.ee/et
Scottish Genomes Partnership (SGP)	The SGP will initially focus on the rapid screening of > 3000 cancer patients in Scotland, diagnosing childhood illnesses, rare genetic diseases and disorders of the central nervous system, also using the data in population studies.	Scottish Government	http://www.scottishgenomespartnership.org/
Danish National Strategy for Personalized Medicine (2017–2020)	The goal is to pave the way for the use of Personalized Medicine in the Danish healthcare system. The first phase was initiated at the beginning of 2017. It will focus on establishing joint governance and a national genome centre. The second phase will be focused on consolidation, research and development.	Danish Government	http://healthcarendenmark.dk/news/new-national-strategy-for-personalized-medicine.aspx
Qatar Genome Programme (QGP)	QGP is in its pilot phase which officially started in September 2015. The initiative aims to map the genome of the local population to apply Precision Medicine in Qatar.	Qatar Government	http://www.qatargenome.org.qa/
Genome of the Netherlands Consortium (GoNL)	GoNL is interested in genetic variation in the Dutch population. To date, the consortium has sequenced 750 whole genomes from Dutch people and 250 trios of two parents and an adult child.	Funded by Netherlands Organization for Scientific Research	http://www.nlgenome.nl/
Australian National Genomic Healthcare Initiative	Australian Genomics is a multi-disciplinary, multi-organization collaboration to integrate genomic medicine into Australian healthcare.	Australia	https://www.australiangenomics.org.au/

The Molecular Analysis for Therapy Choice (NCI-MATCH) and the Molecular Profiling-Based Assignment of Cancer Therapy (NCI-MPACT) trials sponsored by the NCI in the USA are clear examples of precision medicine clinical trials [30–32] (Table 2). In Europe, the PRECISION Panc platform [33] and TRACERx study [34] are ground-breaking efforts to apply precision medicine to clinical trials in the UK. Meanwhile, the Institute Curie is funding the SHIVA trial to evaluate the efficiency of precision medicine approaches, based on the idea that it is as yet unclear whether testing genetic alterations in cancer patients and assigning treatment targeting such alterations is more effective than standard non-targeted therapies. The initial results of SHIVA have shown that molecularly targeted drugs do not improve progression-free survival outside their indications when compared with the treatment of the physician's choice in heavily pre-treated cancer patients [35]. The European Research Council (ERC) has also promoted the Rational molecular Assessments and Innovative Drug Selection (RAIDS) trial, a pioneering effort focused on the development of targeted therapies for cervical cancer [36], as well as the AVATAR trial, where a biopsy of a metastatic lesion from patients will be used to perform a complete exome analysis using next-generation sequencing (NGS). In this trial, a mouse model (Avatar) generated from Patient-Derived Xenografts [37] will be generated for each patient and candidate therapeutic targets can be experimentally tested in the patient's Avatar to select the most effective regimen to ultimately be applied to the patient [38, 39].

Precision bioinformatics for precision medicine

Precision medicine trials will depend on bioinformatics. Indeed, the success of all the aforementioned pioneering studies will largely depend on the strength and quality of the data linking patients, molecular targets and targeted therapies. 'Precision bioinformatics' is defined as a branch of translational bioinformatics that specializes in the development of computational infrastructures, methodologies and tools required for clinical studies that adopt a precision medicine paradigm. Therefore, precision bioinformatics are heavily oriented to: (i) implement platforms for big data processing and integration (including electronic health records—EHRs); (ii) to establish protocols and standards for data exchange that preserve the patients' privacy and that ensure data security; and (iii) to develop computational strategies to exploit, visualize, interpret and summarize such data, ultimately facilitating clinical decision-making [40].

New computational platforms for clinical applications

High-throughput (HTP) technologies and NGS have fuelled the 'Big Data revolution' in biomedicine [41]. For example, the

Table 2. Precision medicine clinical trials

Name	Description	Disease	Location	URL
NCI-MATCH (ID: NCT02465060)	Phase II basket trial designed to study the efficacy of targeted drug treatments tailored directly by genetic testing in advanced solid tumours and lymphomas. The trial will enrol at least 1000 patients in treatment arms based on 19 specific genetic changes, independent of tumour origin.	Solid tumours and lymphomas	Multicentric, USA	https://clinicaltrials.gov/ct2/show/NCT02465060
NCI-MPACT (ID: NCT01827384)	Pilot phase II trial designed to assess if targeted treatments based on precision medicine rational are more effective than standard non-targeted therapies. Molecular profiling-based targeted therapies are prescribed to treat patients with advanced metastatic solid tumours that are usually incurable or not controlled by standard treatments. NCI-MPACT randomly assigns patients with a mutation in a specific genetic pathway to either a targeted therapy for that pathway or a treatment not known to be pathway specific.	Solid tumours	Multicentric, USA	https://clinicaltrials.gov/ct2/show/NCT01827384
PRECISIONPanc	This platform aims to identify more suitable treatment options for pancreatic cancer patients, based on the molecular cancer subtypes defined by their promoters and within the framework of the ICGC consortium. PRECISIONPanc is not yet enrolling patients but it is expected that three clinical trials will recruit around 650 pancreatic cancer patients.	Pancreatic cancer	Multicentric.Sponsor: University of Glasgow, UK	http://www.precisionpanc.org/
TRACERx	TRACERx is analysing the intratumour heterogeneity in approximately 850 stage I–IIIA lung cancer patients and tracking the evolutionary trajectory from diagnosis through to relapse.	Non-small-cell lung cancer	Multicentric.Sponsor: Cancer Research UK	http://www.cruklungcentre.org/Research/TRACERx
SHIVA (ID: NCT01771458)	Randomized phase II trial to assess whether off-label use of commercial drugs for matched molecular alterations confers a clinical benefit to French patients with refractory cancer.	Recurrent metastatic solid tumours	Multicentric.Sponsor: Institut Curie, France.	https://clinicaltrials.gov/show/NCT01771458
RAIDS (ERC ID: 304810)	The RAIDS network collected a prospective data set (BioRAIDs: NCT02428842) of consecutive tumour tissues, whole blood and sera from 419 cervical cancer patients in 2013. At 54 months, whole exome sequencing (WES) became available for the first 98 patients and for 20 CC cell lines, as well as Reverse Phase Protein Array data for 154 patients with a common core set of 91 patients. Targeted sequencing is available for an additional 100 patients. These data allowed the patients to be stratified into different subgroups according to their molecular profile. The correlation with clinical outcomes is currently ongoing.	Cervical cancer	Multicentric.Coordinator: Institut Curie, France.	http://www.raids-fp7.eu/http://cordis.europa.eu/project/rcn/106274_en.html
AVATAR (ERC ID: 670582)	Open label, randomized phase III study of patients receiving standard of care for resistant metastatic pancreatic cancer to test the hypothesis that an integrated personalized approach to treatment improves survival when compared with conventional treatment.	Pancreatic cancer	Multicentric.Madrid, Spain	http://cordis.europa.eu/project/rcn/198713_en.html

Illumina X-Ten System has the capacity to produce around 2 peta-bps per year. This situation now requires precision bioinformatics to implement platforms and algorithms for big data management and integration, ensuring that patient data can be used effectively and that clinically valuable information can be extracted. Diverse initiatives in precision bioinformatics along such lines are described in Table 3. For instance, the need to manage the activities associated with the NCI-MPACT trial motivated the implementation of the GeneMed platform and its public version called OpenGeneMed [42, 43]. The SHIVA and RAIDS trials have implemented the Knowledge and Data Integration (KDI) platform to facilitate data integration, tracking sample processing and delivering a genomic report to facilitate therapeutic decision-making [44]. For its part, the AVATAR trial integrates the RUbioSeq platform to automate sequencing data processing [45] and the PanDrugs method to guide anti-cancer therapy selection (http://www.pandrugs.org/).

Remarkably, a number of local health systems are implementing bioinformatics platforms to support precision medicine studies. For instance, the Geisinger Health System launched DiscovEHR for the longitudinal integration of the EHRs corresponding to the MyCode initiative participants [46, 47]. Moreover, the Personalized Medicine plan was launched in Andalusia (Spain) to develop bioinformatics infrastructures that are interoperable with EHRs. Prototypes have been already tested using a common system for NGS-based diagnostics [48], together with the Medical Genome Project [49], an automated system for the discovery of disease genes [50].

These efforts underlie the need for data integration platforms, together with new computational methodologies and bioinformatics tools to effectively use precision medicine in clinical trials and local healthcare institutions. Other significant projects have also focused on data integration platforms to support precision medicine, such as: the PrecisionFDA platform (https://precision.fda.gov/) for NGS data sharing and pipeline testing; the RD-Connect project (http://rd-connect.eu/) that provides a global and integrated platform in the context of research into rare diseases; the Statistical Multi-Omics Understanding Consortium (SOUND) that focuses on the development of statistical and bioinformatics tools for personal multi-omics data mining (e.g. rDGIdb package [51]); the tranSMART platform to facilitate -omics data exchange [52]; the G-DOC Plus system to combine sequencing data and medical images [53]; and Oncobench [54].

Finally, it is important to highlight that both the 'All of Us' initiative and the European H2020 programme are actively promoting an alignment among public and private sector precision medicine projects [55]. The private sector includes the major pharmaceutical companies, big data, hardware and software companies, and especially, small and medium enterprises. Along these lines, Genomics England has launched the Genomics Expert Network for Enterprises consortium to identify effective and secure ways of bringing industry expertise into the 100 000 Genomes Project [56]. In this scenario, it is reasonable to expect a growing number of public–private partnerships to stimulate the application of commercial bioinformatics platforms in public hospitals and academic institutions. Clear examples include: the IBM Watson Oncology system that has already been introduced into some US cancer centres and hospitals [57]; Microsoft's bot that has been implemented at the Englander Institute of Precision Medicine; the Hortonworks data platform (Table 3); and the WuXi NextCODE alliance that have become associated with both the National Heart Centre in Singapore and the 100 000 Genomes project (https://www.wuxinextcode.com/).

Exchanging patient information to improve healthcare while maintaining privacy

The practical value of developments in precision bioinformatics will depend on continued multidisciplinary discussion involving physicians, data scientists, biostatisticians and experts in clinical bioinformatics, but also, on the adoption of EHRs by health institutes. In the light of this, the Health Information Technology for Economic and Clinical Health provided the US Department of Health and Human Services with $29.5 billion to encourage the adoption of EHRs and to establish meaningful use for interoperability [58].

Beyond the essential digitalization of patient phenotype associated data, it is crucial to establish universal standards and controlled vocabularies to facilitate interoperability and data sharing and to ensure clinical data are collected and interpreted using standardized protocols for privacy and consent. SNOMED CT [59] and the Human Phenotype Ontology [60] represent well-established biomedical ontologies used by the 100 000 Genomes Project. Clear instances of standardization initiatives for clinical informatics protocols include the Clinical Data Interchange Standards Consortium (CDISC; http://www.cdisc.org/standards-and-implementations), Health IT Standards (http://healthcare.nist.gov/), ISO/TC215 (https://www.iso.org/committee/54960.html), Health Information Exchange (https://www.healthit.gov/HIE) and the Health Insurance Portability and Accountability Act (http://www.hhs.gov/ocr/privacy/). Other initiatives focused on general standards for EHRs include Open EHR (http://www.openehr.org/), CDISC (https://www.cdisc.org/standards) and HL7 (http://www.hl7.org/).

Interoperability and integration are crucial yet challenging, particularly given the large heterogeneity of the data collected in EHRs and its confidential nature. In this sense, SemanticHEALTH was a pioneering text mining project focused on gathering fragmented semantic interoperability initiatives from EHRs in the EU [61]. Following SemanticHEALTH, other efforts have recently been initiated, and SemanticHealthNet (http://www.semantichealthnet.eu/) was launched to develop a scalable and sustainable pan-European semantic interoperable protocol for clinical and biomedical knowledge, and to help ensure that EHR systems are optimized for patient care, public health and clinical research in different healthcare systems and institutions [62]. Other collaborative initiatives for the systematic integration of EHRs include: P-medicine (http://www.p-medicine.eu/), focused on developing secure tools, robust data sharing and integration systems, IT infrastructure and virtual physiological human models to support precision medicine [63, 64]; and CER Hub, a web-based platform to combine comprehensive electronic clinical data from multiple healthcare organizations [65].

In addition to the heterogeneity of the data and the poor standardization, there are other technical and non-technical bottlenecks that must be overcome to establish the sustainable exchange of information via EHRs and genomic studies that is necessary for precision medicine. These include a lack of incentives, specifically in relation to the loss of patients to other hospitals. This problem can be overcome by making patient's health data available anywhere, although it has long been recognized that healthcare systems are often reticent to participate in health information exchanges. Other recognized difficulties include the inefficient sorting through excessively non-selective patient information and problems in understanding the data shared in a medical context, particularly owing to a lack of details associated with the clinical notes driven by

Table 3. Precision bioinformatics infrastructure initiatives

Initiative	Description	Promoter	Location	URL
GeneMed	GeneMed is an informatics system designed to favour collaboration between a sequencing lab, the treatment selection team and clinical personnel, to reduce errors made by transferring and sharing data between groups, and to aid clear documentation in the NCI-MPACT clinical trial pipeline.	NCI-MPACT	USA	—
OpenGeneMed	Public version of the GeneMed system.	NCI-MPACT	USA	https://brb.nci.nih.gov/OpenGeneMed/
DiscovEHR	The DiscovEHR browser facilitates access to variant frequency data from >50 000 MyCode participants. It facilitates allele frequency comparisons with other population-based and biobank resources.	Regeneron Genetics Center and Geisinger Health System MyCode	USA	http://www.discovehrshare.com/
KDI	Informatics platform implemented to ensure information sharing, cross-software interoperability, automatic data extraction and secure data transfer in the context of SHIVA, RAIDS and other studies. KDI is currently used to manage all the high-throughput data at the Institute Curie.	Institute Curie	France	—
PrecisionFDA	A cloud-based data sharing system to evaluate NGS assays and for regulatory science exploration.	FDA	USA	https://precision.fda.gov/
RD-Connect	An infrastructure project that brings together databases, registries, biobanks and clinical bioinformatics data used for rare disease research in a central resource for researchers worldwide.	EU	Europe	http://rd-connect.eu/
SOUND	An international consortium established to create bioinformatics tools for statistically informed use of personal genomic and other 'omics data in a medical context.	EU	Europe	http://www.sound-biomed.eu/
tranSMART	An open-source cloud system implemented to facilitate 'omics data exchange in clinical and translational research.	tranSMART foundation	USA-Europe	http://transmartfoundation.org/
G-DOC Plus	Data integration and bioinformatics cloud platform to handle diverse biomedical big data, including gene expression arrays, NGS and medical images.	Georgetown University	USA	https://gdoc.georgetown.edu/gdoc/
Oncobench	A platform developed to analyse tumour data in clinical practice. The current version is collecting up to 0.5 Gb of DNA data per patient.	Geneva University Hospitals, the Swiss Institute of Bioinformatics and others.	Switzerland	—
Watson Oncology	A cognitive computing system designed to support clinical decision making and to interpret cancer patients' clinical information, identifying individualized, evidence-based treatment options.	IBM and several US hospitals.	USA	https://www.mskcc.org/about/innovative-collaborations/watson-oncology
Precision Medicine Knowledgebase Bot	PMKB Bot connects to several channels, including Microsoft Teams, Skype, Slack and WebChat. As a result, clinicians can access these data in many different ways and make clinical decisions faster. PMKB currently supports 163 genes and 518 variants with 404 clinical interpretations.	Microsoft and Englander Institute for Precision Medicine.	USA	https://pmkb.weill.cornell.edu/
Hortonworks Data Platform (HDP)	A platform to store and process huge amounts of liver cancer data, making that data and related tools accessible to researchers in five different teams. HDP cluster at Arizona State University has accumulated more than a petabyte of genomic data from multiple studies involving over 500 individuals in each.	Hortonworks and Arizona State University.	USA	https://es.hortonworks.com/customers/arizona-state-university/

privacy concerns [66]. The reticence of patients and providers to exchange information based on privacy concerns and reports of healthcare data breaches is not unfounded [67]. However, the GA4GH Beacon Project (https://beacon-network.org/), which responds to allele-presence queries, represents an attempt to share genomic data without revealing individual patient information. However, this limits the utility of the data for research and diagnosis, and it is not exempt from risks of patient

re-identification attacks [68]. The conflict between the need for data sharing and the maintenance of data security has also stimulated active research into novel cryptographic implementations that serve both patient healthcare, and that protect sensitive genomic information and patient privacy [69, 70].

It is also necessary to note that precision medicine studies should describe the full bioinformatics settings used to evaluate the quality and the traceability of the data generated. Apart from the aforementioned ClinVar resource, used regularly by clinical geneticists and other clinicians worldwide, the SHIVA trial has set a good precedent in this sense. In particular, this clinical trial fully reported the ontologies used, together with the standards and the complete bioinformatics framework used in its pipeline [44]. This kind of good practice promotes transparency, it serves as a reference for other studies and it facilitates pipeline reproducibility to those who will be in charge of computational analysis and multi-omics data interpretation in precision medicine studies (i.e. the clinical bioinformaticians).

Box 1. Fundamental technical skills for clinical bioinformaticians

1. Informatics
○ Experience in UNIX command line.
○ A basic programming language (i.e. Python). R as a useful language for handling statistics.
○ Knowledge of big data environments.
2. Life sciences
○ Understand the different types of biological data and databases.
○ Comprehend HTP data analysis methods.
○ Multi-omics data integration and interpretation.
3. Clinical scenario
○ Be familiar with EHRs, clinical terminology and medical procedures and protocols.
○ Get to know medical genomics: diagnosis, predictive and prognosis biomarkers.
○ Understand clinical trial design and monitoring.

Clinical bioinformaticians at hospitals: the pioneer comes to town

As seen above, the challenges to successfully apply precision medicine in hospitals are numerous and in particular, they are related to the need for large computational infrastructures for data processing and storage; the establishment of guidelines, standards and controlled protocols; the implementation of security policies to access data, along with biobanking, legal issues and ethical questions; and, last but not least, the lack of trained and specialized professionals to perform such tasks.

At first glance, this situation resembles what happened two decades ago in genomics, where the completion of the Human Genome project and the development of multi-omics HTP methods highlighted the need for trained bioinformaticians to handle, analyse and extract information from large amounts of data [71]. Indeed, with the current big data explosion, this situation is even more evident, and consequently, experienced bioinformaticians and data scientist are in high demand by the big pharma and biotech companies, and at academic institutions [72]. The most desired professional profiles include experts at the crossroads of computer sciences, statistics and molecular biology, those able to present big data information in a clear manner to decision-makers [73, 74]. Nevertheless, such a profile might not suffice to cover all precision medicine requirements, which require an expert with bioinformatics skills who is familiar with the hospital environment and its specific particularities, who is fluent in clinical terminology, acquainted with clinical trial design and procedures, and most importantly, capable of understanding physicians demands for clinical decision-making tools as a priority to improve the standard of patient care (Box 1). This type of bioinformatician, orientated towards the clinical environment, is still rare, yet it is an increasingly demanded species in today's hospitals. Encountering such specialists is imperative for the effective implementation of precision medicine in health institutions, which will certainly need to set up specialist teams of clinical bioinformaticians to bridge the gap between the patient's genomic landscape and clinical decision-making, linking raw sequences, data analysis tools, interpretation algorithms and a variety of databases and EHR repositories for each particular clinical case (Figure 2).

Alternatively, it is clear that the expansion of precision medicine in the clinical environment will require an effort from physicians to better understand how bioinformatics can help in their work, assisting data-driven clinical decision-making. In fact, a number of authors have remarked on this issue, claiming physicians should become knowledgeable users who can understand the output from the bioinformatics analysis of a patient's genomic data [75–78]. Indeed, other recent proposals go beyond specialized training in computational biology for physicians, proposing the integration of translational bioinformatics into the medical curriculum, and the establishment of official precision medicine certificates for clinical geneticists and health professionals interested in this field [79, 80]. Although the natural targets of these proposals are first clinical geneticists, these professionals are aware of their lack of training in bioinformatics, and the working duo of geneticist–bioinformatician is becoming a more common finding in hospitals. The landscape is not that clear for other medical specialties such as Oncology, Cardiology or Neurology, where the bioinformatician's competences are considered far removed from the daily clinical work. Consequently, it is not surprising that a wide variety of opportunities for medical staff to train in bioinformatics have emerged in recent years [81]. Such opportunities must ensure they focus both on knowledge acquisition and clear practical examples of clinical use, more directly orientated to 'classical' health professionals.

At present it is hard to find proposals for clinical training that specifically target bioinformaticians and computational biologists. The absence of such training in the bioinformatics community is slowing down the integration of bioinformaticians into the healthcare sector, contributing to a deceleration in the implementation of precision medicine at healthcare institutions. Undoubtedly, bioinformaticians and physicians speak different languages and have distinct scientific cultures that are not always easy to reconcile. The clinical bioinformatics community should contribute to overcome this language barrier, adapting its curriculum to a medical scenario and adding clinical knowledge to its background to facilitate communication with health professionals. It is necessary to organize and promote courses in fundamental clinical areas for bioinformatics professionals to ensure their full integration into working teams in healthcare institutions (Figure 3). Precision medicine will definitely benefit from the incorporation of clinically trained

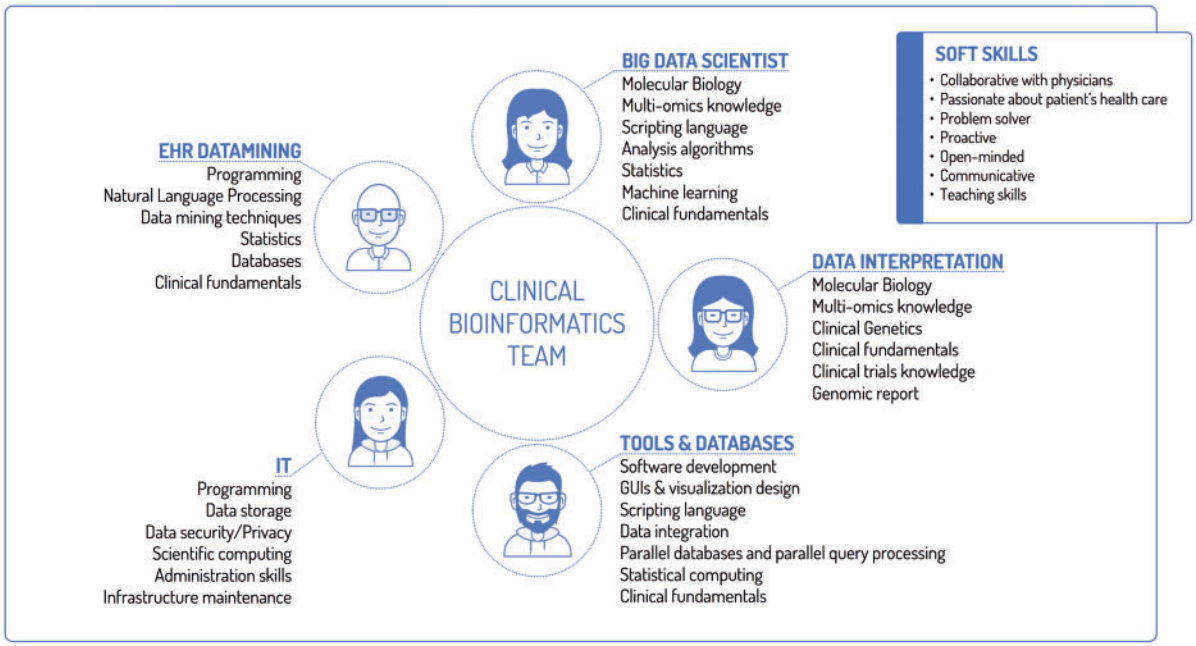

Figure 2. Clinical bioinformatics laboratory profile. Clinical bioinformatics teams require multidisciplinary experts to perform regular tasks. Computational experts will be in charge of servers, databases, development and pipeline optimization. More biologically focused profiles will be responsible for genomic analysis and interpretation. In addition, all team members will have been trained to a varying degree in basic clinical sciences. There will be a continued knowledge exchange among the team members to achieve clinical goals and improve patient healthcare.

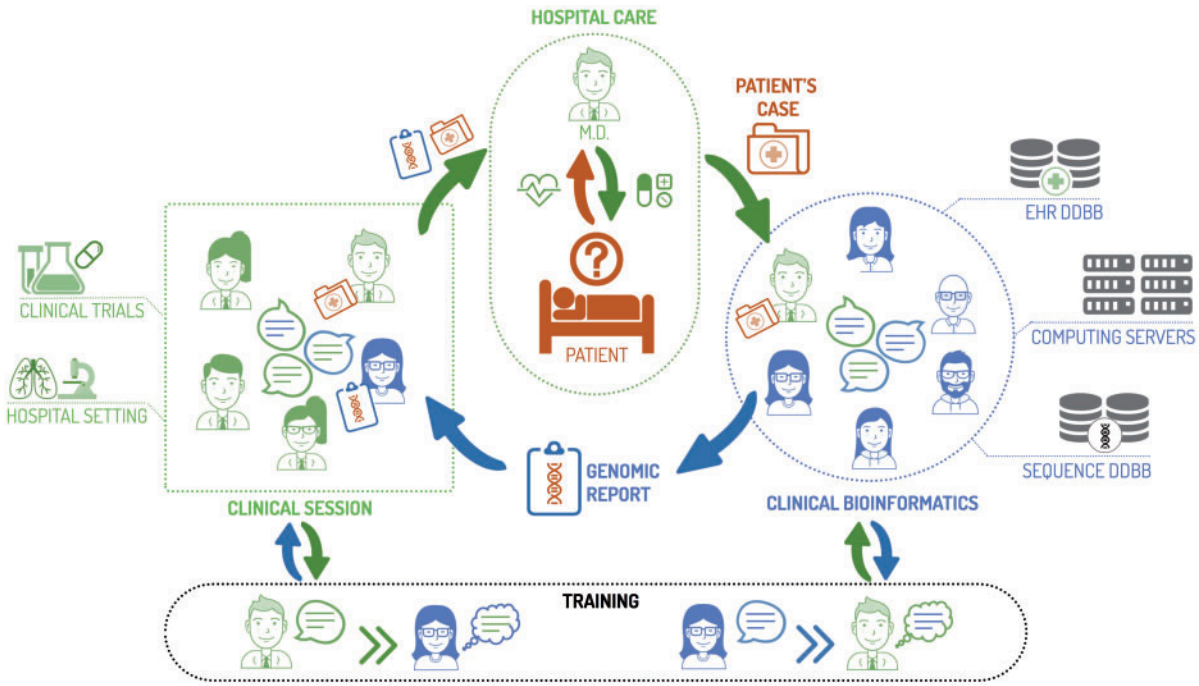

Figure 3. Precision medicine workflow in hospitals. The patient's standard of care in a precision medicine scenario requires specialist clinical bioinformaticians to participate in multidisciplinary clinical sessions with physicians and clinical geneticists, making knowledge exchange easier and facilitating data-driven diagnosis and clinical decision-making. Conversely, physicians and clinical geneticists should be incorporated into clinical bioinformatics discussions to guide and adapt genomic reports to the medical reality. Clinical bioinformaticians, in hospitals should not be simple data extraction technicians but rather, specialized partners of physicians in accordance with a patient-centred model. Mutual training for clinicians and bioinformaticians should provide reciprocal and bidirectional information exchange.

bioinformaticians capable of straddling the medical, biological and computational worlds, who can write code, interpret genome data and communicate with health professionals in clinical departments and hospitals.

Approaching physicians: emerging efforts in clinical bioinformatics training

In Europe, the ELIXIR Training Platform represents a solid training initiative that offers guidelines and best practices for

educational excellence in bioinformatics, focused on the training of bioinformaticians and computational biologists [82, 83]. In addition, the Global Organisation for Bioinformatics Learning, Education and Training (GOBLET; http://mygoblet.org) and the International Society for Computational Biology Education Committee offer a networking structure for bioinformatics trainers and trainees [84]. Although the ELIXIR Training Platform and GOBLET have traditionally been involved in teaching bioinformatics for life sciences, the current demands for clinical bioinformaticians have boosted new training efforts to overcome this bottleneck. Remarkably, the ELIXIR-EXCELERATE project and UK's Health Education England (HEE) have recently set up training workshops on clinical bioinformatics in the UK, focusing on bioinformatics instructors, and providing guidance and tips to develop and deliver training in clinical bioinformatics [85]. Moreover, the urgent need for early training in clinical bioinformatics skills drove the HEE team to elaborate a report that sets out a phased approach to provide the training required to support the 100 000 Genomes Project within the UK healthcare system. This report claims that the first urgent need for precision medicine and medical genomics is to recruit specialist healthcare scientists to assist clinical staff with the interpretation of genome sequencing data. This includes a detailed long-term training plan that promotes PhD and postdoctoral fellowships in clinical bioinformatics, as well as MSc programmes in Medical Genomics, to develop and integrate workforces of bioinformaticians into British healthcare institutions [86].

This HEE strategy emphasizes the idea that implementing precision medicine requires a novel bioinformatics expert able to integrate and interact with the hospital environment. To do so, clinical bioinformaticians must be trained in basic clinical principles, clinical trials procedures and medical genomics. Such a specific clinical background would facilitate efficient and natural communication with physicians, medical geneticists and other healthcare specialists. Clinical training would also help bioinformaticians realize that their role in hospitals goes beyond data extraction, storage and interpretation. As new partners to physicians, clinical bioinformaticians should actively participate in clinical sessions, supporting clinical decision-making with clear and intuitive genomic reports designed on the basis of input from the clinical staff (Figure 3). Clinical courses, workshops and educational programmes orientated towards bioinformaticians, along with reciprocal training for healthcare specialists, would help to increase fluent and effective information transfer for the ultimate benefit of patient care.

Box 2. Tips for emergent clinical bioinformatics group leaders

1) Rule of thumb: the patient's health is at the centre of everything you do.
2) You'll need complementary roles in your team, as computational, biological and clinical profiles are interdependent.
3) Be pragmatic, the ultimate goal is to fulfil medical needs to facilitate clinical decision-making.
4) Stay up-to-date, parallel research projects are essential to achieve excellence in genomics service.
5) Be rigorous (yes, even more rigorous), especially as you will be dealing with patient data, not just research data. Your analysis will contribute to the patient's healthcare.
6) Open-mindedness and flexibility are essential. In hospitals, hardly anyone shares their bioinformatics background. Whoever adapts, wins.
7) Be communicative, overcome your own scientific limitations and language barriers.
8) Hone your teaching skills, as you will need to explain your protocols and results repeatedly.
9) Do not isolate yourself. Commands and computational papers are ok but, well... you work with healthcare professionals and patients. There is a world out there beyond computers and algorithms.
10) You are not alone. Networking with bioinformatics colleagues in other health institutions really works. Share your knowledge...and your ignorance.

Perspectives: the evolution of precision bioinformatics

The molecular diagnostics laboratories of tomorrow will depend on trained bioinformaticians and until this happens, precision medicine will drive for the incorporation of an increasing number of specialist bioinformaticians at healthcare institutions. It is expected that these diagnostic laboratories will offer sequencing services to healthcare systems and create efficient informatics solutions to execute such tasks, all under an approved regulatory process. Pioneering groups are currently incorporating bioinformaticians, albeit with limited experience in clinical procedures and hospital environments (Box 2). Clearly, the integration of these novel specialists is not a trivial task. Indeed, there are a number of technical and non-technical barriers that need to be addressed and these will first require a strong commitment and decisive efforts by the institutions and their clinical staff, but also from the whole bioinformatics community.

One such technical barrier is related to the high-performance computing infrastructures required for the efficient storage, processing and interpretation of routine large-scale genomic analysis in national healthcare systems. Unfortunately, the computing infrastructures that are currently found at healthcare institutions are not usually prepared to efficiently process such volumes of data. To address this issue in the most cost-effective way, and to deliver a genomics-based service to the British health service, Genome England is considering the possibility of establishing a pilot semi-centralized system of molecular diagnostics laboratories from 10 to 20 British cancer centres. These laboratories would share standardized operating procedures and NGS platforms. Raw sequence and phenotype data would be stored at each clinical site locally, yet data analysis would be carried out in a single cloud environment using a standardized mutation detection pipeline [87]. Other European strategies, such as EuroHPC [88] and the Partnership for Advanced Computing in Europe, are working towards the establishment of a multi-government cooperative framework to acquire and deploy an integrated supercomputing infrastructure. It is expected that EuroHPC will develop a test bed to create large European digital infrastructures for personalized medicine data.

Such approaches, based on cloud computing, have been successfully applied in massive sequencing projects like the PanCancer Analysis Whole Genome project, where full genome analyses were accomplished in a cloud-computer-based

architecture across 13 data centres distributed over three continents. However, these strategies need to be studied thoroughly to be applicable in clinical settings. In fact, some authors reported concerns regarding the utilization of cloud-based systems for patient data computing. These concerns are related to perceived limitations in data security and protection, the need for due consideration of the rights of patient donors and research participants, and legal issues associated to local regulations owing to fundamental differences in the understanding of the right to data protection between different legal systems [89].

Without a doubt, the implementation of shared supercomputing initiatives to support precision medicine data processing will require a substantial economic investment on the part of governments. Nevertheless, sequencing costs are falling and it is expected that the creation of shared computing infrastructures to support sequencing on a larger scale will help push sequencing costs down further, with consequent savings in time and money for healthcare systems. In addition, it is expected that precision medicine will eventually help healthcare institutions save money in treatments, while at the same time enhancing the quality of life of patients. For instance, whole-exome sequencing (WES) of children's rare disease was shown to have improved the diagnostic rate 5-fold compared with standard care, while reducing costs [90]. Similarly, other studies indicate that the use of WES as an early, routine clinical test for infants with suspected monogenic disorders more than triples the successful diagnosis, at one-third of the cost [91]. These are clear examples of the application of panels and exome sequencing already implemented in molecular diagnostics laboratories at hospitals. These alternatives to WGS have a significantly smaller processing burden and they are therefore cheaper, although they cannot compete with the capacity of WGS to offer a complete genome landscape and more information about structural variations. It is also expected that WGS will be industrialized as a single common process, becoming more cost-effective than panels or exomes. Accordingly, a patient's whole genome data will be used as a multi-diagnostic test for a catalogue of diseases or clinical phenotypes [87].

The integration and exploitation of population-scale biomedical data generated by mobile and real-time wellness monitoring devices also raises significant challenges for precision medicine bioinformatics. It is expected that monitorized data would promote data-driven public health studies to design precision prevention strategies [16, 92]. Such data would also help to stratify patients for active health management, dealing better with clinically asymptomatic patients and their underlying medical history. Clinical bioinformatics tools and resources are fundamental for these advances in implementing real-time biomedical and healthcare analyses in a clinical environment. The development of data capturing and storage strategies that integrate and correlate health data with patient's EHRs on a population scale is an issue that must be addressed, and it will require the implementation of novel scientific and technical resources and methods [16]. Another forthcoming hurdle is related to radiomics, an emerging and promising field that involves the HTP extraction of many features from radiographic images [93]. Pioneering radiomics approaches have recently been incorporated into precision oncology, whereby tumour characterization is not just limited to anatomy but it can also reveal information at the cellular and genomic level that can be quantified as an imaging phenotype. Thus, computational algorithms for radiomics allow quantitative automated imaging features to be converted into mineable data. Pioneering studies in head and neck [94] or lung cancer [95] provide preliminary evidence that radiomics texture analyses can define distinctive tumour phenotypes that are driven by underlying genotypes. In this way, radiomics signatures can hold predictive and prognostic information to guide personalized radiotherapy. It is expected that the development of computational methods to efficiently extract and process radiomics data, in conjunction with the implementation of platforms to readily integrate radiomics with clinical, pathological and genomic information, will be research areas of great interest in coming years.

In summary, the emerging precision medicine paradigm offers a new and challenging scenario to generate an unprecedented amount of health-related data. A novel type of bioinformatician is required to translate such data into knowledge that can be used to facilitate clinical decision-making. Clinical bioinformaticians will unify their efforts with the staff at hospitals and healthcare institutions to successfully deliver quality patient care. This situation provides new, attractive and challenging job opportunities to bioinformaticians, and maybe it is time to ask not what their health institutions can do for them but rather, what they can do for their health institutions.

Key Points

- The precision medicine initiatives emerging around the world are facing many challenges.
- Current public and private investments aim to establish the computational infrastructures required to support precision medicine initiatives.
- Precision medicine and clinical bioinformatics can only work efficiently if electronic health records and patient genotypes are accessible to in-house bioinformaticians.
- The successful implementation of precision medicine in health institutions requires bioinformaticians with a basic clinical training, as yet an unfulfilled need.
- The clinical bioinformatician is a novel and specialized profile demanded increasingly by healthcare centres.

Acknowledgements

Authors would like to thank all members from CNIO Bioinformatics Unit for their comments, all of which have helped improve this manuscript. A.V. to BSC-IRB-CRG Program in Computational Biology and Award Severo Ochoa, SEV 2015-0493.

Funding

This study was funded by the following grants: BIO2014-57291-R from the Spanish Ministry of Economy and Competitiveness and PT13/0001/0007 from the ISCIII, both co-funded with European Regional Development Funds (ERDF); EU H2020-INFRADEV-1-2015-1 ELIXIR-EXCELERATE (ref. 676559); and Marie-Curie Career Integration Grant (CIG) CIG334361.

References

1. Biankin AV. The road to precision oncology. *Nat Genet* 2017; **49**(3):320–1.
2. Duffy DJ. Problems, challenges and promises: perspectives on precision medicine. *Brief Bioinform* 2016;**17**(3):494–504.
3. Valencia A, Hidalgo M. Getting personalized cancer genome analysis into the clinic: the challenges in bioinformatics. *Genome Med* 2012;**13**(7):61.
4. Dubitzky W. Computational systems biomedicine. *Brief Bioinform* 2016;**17**(3):367.

5. Vogelstein B, Papadopoulos N, Velculescu VE, *et al*. Cancer genome landscapes. *Science* 2013;**339**(6127):1546–58.
6. Lawrence MS, Stojanov P, Polak P, *et al*. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 2013;**499**(7457):214–8.
7. Gerlinger M, Rowan AJ, Horswell S, *et al*. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012;**366**(10):883–92.
8. Rubio-Perez C, Tamborero D, Schroeder MP, *et al*. *In silico* prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. *Cancer Cell* 2015;**27**(3):382–96.
9. Alizadeh AA, Aranda V, Bardelli A, *et al*. Toward understanding and exploiting tumor heterogeneity. *Nat Med* 2015;**21**(8):846–53.
10. Eifert C, Powers RS. From cancer genomes to oncogenic drivers, tumors dependencies and therapeutic targets. *Nature* 2012;**12**:572–6.
11. Hyman DM, Taylor BS, Baselga J. Implementing genome-driven oncology. *Cell* 2017;**168**(4):584–99.
12. Redig AJ, Jänne PA. Basket trials and the evolution of clinical trial design in an era of genomic medicine. *J Clin Oncol* 2015;**33**(9):975–7.
13. Kandoth C, McLellan MD, Vandin F, *et al*. Mutational landscape and significance across 12 major cancer types. *Nature* 2013;**502**(7471):333–9.
14. Clinical Cancer Genome Task Team of the Global Alliance for Genomics and Health. Sharing clinical and genomic data on cancer—the need for global solutions. *N Engl J Med* 2017;**376**:2006–9.
15. Lawler M, Maughan T. From rosalind franklin to barack obama: data sharing challenges and solutions in genomics and personalised medicine. *New Bioeth* 2017;**23**(1):64–73.
16. Shameer K, Badgeley MA, Miotto R, *et al*. Translational bioinformatics in the era of real-time biomedical, healthcare and wellness data streams. *Brief Bioinform* 2017;**18**(1):105–24.
17. Longo DL, Drazen JM. Data sharing. *N Engl J Med* 2016;**374**(3):276–7.
18. Bierer BE, Crosas M, Pierce HH. Data authorship as an incentive to data sharing. *N Engl J Med* 2017;**376**(17):1684–7.
19. Rosenbaum L. Bridging the data-sharing divide—seeing the devil in the details, not the other camp. *N Engl J Med* 2017;**376**(23):2201–3.
20. Reardon S. Giant study poses DNA data-sharing dilemma. *Nature* 2015;**525**(7567):16–7.
21. Vis DJ, Lewin J, Liao RG, *et al*. Towards a global cancer knowledge network: dissecting the current international cancer genomic sequencing landscape. *Ann Oncol* 2017;**28**(5):1145–51.
22. Feldman EA. The Genetic Information Nondiscrimination Act (GINA): public policy and medical practice in the age of personalized medicine. *J Gen Intern Med* 2012;**27**(6):743–6.
23. Dzau VJ, Geoffrey SG. Realizing the full potential of precision medicine in health and health care. *Jama* 2016;**316**(16):1659–60.
24. AACR Project GENIE Consortium. AACR project GENIE: powering precision medicine through an international consortium. *Cancer Discov* 2017;**7**(8):818–31.
25. Global Alliance for Genomics and Health. GENOMICS. A federated ecosystem for sharing genomic, clinical data. *Science* 2016;**352**(6291):1278–80.
26. http://genomicsandhealth.org/work-products-demonstration-projects/catalogue-global-activities-international-genomic-data-initiati.
27. Philippakis AA, Azzariti DR, Beltran S, *et al*. The matchmaker exchange: a platform for rare disease gene discovery. *Hum Mutat* 2015;**36**(10):915–21.
28. Landrum MJ, Lee JM, Benson M, *et al*. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* 2016;**44**(D1):D862–8.
29. Grossman RL, Heath AP, Ferretti V, *et al*. Toward a shared vision for cancer genomic data. *N Engl J Med* 2016;**375**(12):1109–12.
30. Coyne GO, Takebe N, Chen AP. Defining precision: the precision medicine initiative trials NCI-MPACT and NCI-MATCH. *Curr Probl Cancer* 2017;**41**(3):182–93.
31. Brower V. NCI-MATCH pairs tumor mutations with matching drugs. *Nat Biotechnol* 2015;**33**(8):790–1.
32. Lih CJ, Sims DJ, Harrington RD, *et al*. Analytical validation and application of a targeted next-generation sequencing mutation-detection assay for use in treatment assignment in the NCI-MPACT trial. *J Mol Diagn* 2016;**18**(1):51–67.
33. Bailey P, Chang DK, Nones K, *et al*. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* 2016;**531**(7592):47–52.
34. Jamal-Hanjani M, Hackshaw A, Ngai Y, *et al*. Tracking genomic cancer evolution for precision medicine: the lung TRACERx study. *PLoS Biol* 2014;**12**:e1001906.
35. Le Tourneau C, Delord JP, Gonçalves A, *et al*. Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): a multicentre, open-label, proof-of-concept, randomised, controlled phase 2 trial. *Lancet Oncol* 2015;**16**(13):1324–34.
36. Samuels S, Balint B, von der Leyen H, *et al*. Precision medicine in cancer: challenges and recommendations from an EU-funded cervical cancer biobanking study. *Br J Cancer* 2016 6;**115**(12):1575–83.
37. Hidalgo M, Amant F, Biankin AV, *et al*. Patient-derived xenograft models: an emerging platform for translational cancer research. *Cancer Discov* 2014;**4**(9):998–1013.
38. Garralda E, Paz K, López-Casas PP, *et al*. Integrated next-generation sequencing and avatar mouse models for personalized cancer treatment. *Clin Cancer Res* 2014;**20**(9):2476–84.
39. Byrne AT, Alférez DG, Amant F, *et al*. Interrogating open issues in cancer precision medicine with patient-derived xenografts. *Nat Rev Cancer* 2017;**17**(4):254–68.
40. Rehm HL. Evolving health care through personal genomics. *Nat Rev Genet* 2017;**18**(4):259–67.
41. Eisenstein M. Big data: the power of petabytes. *Nature* 2015;**527**(7576):S2–4.
42. Zhao Y, Polley EC, Li MC, *et al*. GeneMed: an informatics hub for the coordination of next-generation sequencing studies that support precision oncology clinical trials. *Cancer Inform* 2015;**14**:45–55.
43. Palmisano A, Zhao Y, Li MC, *et al*. OpenGeneMed: a portable, flexible and customizable informatics hub for the coordination of next-generation sequencing studies in support of precision medicine trials. *Brief Bioinform* 2017;**18**:723–34. doi: 10.1093/bib/bbw059.
44. Servant N, Roméjon J, Gestraud P, *et al*. Bioinformatics for precision medicine in oncology: principles and application to the SHIVA clinical trial. *Front Genet* 2014;**5**:152.
45. Rubio-Camarillo M, Gómez-López G, Fernández JM, *et al*. RUbioSeq: a suite of parallelized pipelines to automate exome variation and bisulfite-seq analyses. *Bioinformatics* 2013;**29**(13):1687–9.
46. Carey DJ, Fetterolf SN, Davis FD, *et al*. The Geisinger MyCode community health initiative: an electronic health record–linked biobank for precision medicine research. *Genet Med* 2016;**18**(9):906–13.
47. Dewey FE, Murray MF, Overton JD, *et al*. Distribution and clinical impact of functional variants in 50, 726 whole-exome sequences from the DiscovEHR study. *Science* 2016;**354**.
48. Alemán A, Garcia-Garcia F, Medina I, *et al*. A web tool for the design and management of panels of genes for targeted enrichment and massive sequencing for clinical applications. *Nucleic Acids Res* 2014;**42**(W1):83–7.

49. Dopazo J, Amadoz A, Bleda M, *et al*. 267 Spanish exomes reveal population-specific differences in disease-related genetic variation. *Mol Biol Evol* 2016;**33**(5):1205–18.
50. Alemán A, Garcia-Garcia F, Salavert F, *et al*. A web-based interactive framework to assist in the prioritization of disease candidate genes in whole-exome sequencing studies. *Nucleic Acids Res* 2014;**42**(W1):88–93.
51. Thurnherr T, Singer F, Stekhoven DJ, *et al*. Genomic variant annotation workflow for clinical applications. *F1000Res* 2016;**5**:1963.
52. Bauer CR, Knecht C, Fretter C, *et al*. Interdisciplinary approach towards a systems medicine toolbox using the example of inflammatory diseases. *Brief Bioinform* 2017;**18**:479–87.
53. Bhuvaneshwar K, Belouali A, Singh V, *et al*. G-DOC Plus—an integrative bioinformatics platform for precision medicine. *BMC Bioinformatics* 2016;**17**(1):193.
54. http://www.sib.swiss/about-us/news/news-2016/966-launch-of-a-new-cancer-diagnostic-platform-oncobench
55. Granados Moreno P, Joly Y, Knoppers BM. Public–private partnerships in cloud-computing services in the context of genomic research. *Front Med* 2017;**4**:3.
56. https://www.genomicsengland.co.uk/clinicians-researchers-and-industry-collaborate-with-the-100000-genomes-project/
57. Bringing precision medicine to community oncologists. *Cancer Discov* 2017;**7**:6–7.
58. Charles D, King J, Patel V, *et al*. *Adoption of Electronic Health Record Systems among U.S. Non-federal Acute Care Hospitals: 2008–2012. ONC Data Brief, no 9*. Washington, DC: Office of the National Coordinator for Health Information Technology, 2013.
59. Millar J. The need for a global language—SNOMED CT introduction. *Stud Health Technol Inform* 2016;**225**:683–5.
60. Robinson PN, Mundlos S. The human phenotype ontology. *Clin Genet* 2010;**77**(6):525–34.
61. Stroetman VN, Kalra D, Lewalle P, *et al*. *Semantic Interoperability for Better Health and Safer Healthcare Research and Deployment Roadmap for Europe*. Directorate-General for the Information Society and Media (European Commission), 2009. DOI: 10.2759/38514. https://publications.europa.eu/en/publication-detail/-/publication/9bb4f083-ac9d-47f8-ab4a-76a1f095ef15/language-en
62. Legaz-García Mdel C, Martínez-Costa C, Miñarro-Giménez JA, *et al*. Ontology patterns-based transformation of clinical information. *Stud Health Technol Inform* 2014;**205**:1018–22.
63. Schera F, Weiler G, Neri E, Kiefer S, *et al*. The p-medicine portal-a collaboration platform for research in personalised medicine. *Ecancermedicalscience* 2014;**11**:398.
64. Marés J, Shamardin L, Weiler G, *et al*. p-medicine: a medical informatics platform for integrated large scale heterogeneous patient data. *AMIA Annu Symp Proc* 2014;**14**:872–81.
65. Hazlehurst BL, Kurtz SE, Masica A, *et al*. CER Hub: an informatics platform for conducting comparative effectiveness research using multi-institutional, heterogeneous, electronic clinical data. *Int J Med Inform* 2015;**84**(10):763–73.
66. Sitpati A, Kim H, Berkovich B, *et al*. Integrated precision medicine: the role of electronic health records in delivering personalized treatment. *WIREs Syst Biol Med* 2017;**9**:e1378-12.
67. http://www.hipaajournal.com/largest-healthcare-data-breaches-of-2016-8631/.
68. Raisaro JL, Tramèr F, Ji Z, *et al*. Addressing Beacon re-identification attacks: quantification and mitigation of privacy risks. *J Am Med Inform Assoc* 2017;**14**:799–805. doi: 10.1093/jamia/ocw167
69. Jagadeesh KA, Wu DJ, Birgmeier JA, *et al*. Deriving genomic diagnoses without revealing patient genomes. *Science* 2017;**357**(6352):692–5.
70. Simmons S, Sahinalp C, Berger B. Enabling privacy-preserving GWASs in heterogeneous human populations. *Cell Syst* 2016;**3**(1):54–61.
71. Wickware P. Training in a hybrid discipline. *Nature* 2001;**413**(6858):6–7.
72. Jeffrey C. Core services: reward bioinformaticians. *Nature* 2015;**520**:151–2.
73. Levine AG. An explosion of bioinformatics careers. *Science* 2014;**344**:1303–06.
74. Spotlight on bioinformatics. Biology goes digital. *NatureJobs* 2016. doi: 10.1038/nj0478.
75. Lopez-Campos G, Lopez-Alonso V, Martin-Sanchez F. Training health professionals in bioinformatics. Experiences and lessons learned. *Methods Inf Med* 2010;**49**(3):299–304.
76. Rubinstein JC. Perspectives on an education in computational biology and medicine. *Yale J Biol Med* 2012;**85**(3):331–7.
77. Brazas MD, Lewitter F, Schneider MV, *et al*. A quick guide to genomics and bioinformatics training for clinical and public audiences. *PLoS Comput Biol* 2014;**10**(4):e1003510.
78. Clay MR, Fisher KE. Bioinformatics education in pathology training: current scope and future direction. *Cancer Inform* 2017;**10**:16.
79. Tan B, Ban KH, Tan TW. Integrating translational bioinformatics into the medical curriculum. *Int J Med Educ* 2014;**5**:132–4.
80. McGrath S, Ghersi D. Building towards precision medicine: empowering medical professionals for the next revolution. *BMC Medi Genomics* 2016;**9**(1):23.
81. Rozman D, Acimovic J, Schmeck B. Training in systems approaches for the next generation of life scientists and medical doctors. *Methods Mol Biol* 2016;**1386**:73–86.
82. Via A, Blicher T, Bongcam-Rudloff E, *et al*. Best practices in bioinformatics training for life scientists. *Brief Bioinform* 2013;**14**(5):528–37.
83. Tramontano A, Valencia A. Education and research infrastructures. In: A. Cesario, F.B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine*. Springer Science, Business Media B.V., 2011, pp. 165–181. DOI: 10.1007/978-94-007-1567-7_6. ISBN 978-94-007-1567-7.
84. Attwood TK, Bongcam-Rudloff E, *et al*. GOBLET, the global organisation for bioinformatics learning, education and training. *Nature* 2017;**544**:e1004143.
85. https://www.elixir-europe.org/events/excelerate-train-trainer-clinical-bioinformatics-and-best-practices-workshop
86. https://www.genomicseducation.hee.nhs.uk/images/publications/Developing_NHS_Clinical_Bioinformatics_Training.pdf
87. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/624628/CMO_annual_report_generation_genome.pdf
88. https://ec.europa.eu/digital-single-market/en/news/eu-ministers-commit-digitising-europe-high-performance-computing-power
89. Molnár-Gábor F, Lueck R, Yakneen S, *et al*. Computing patient data in the cloud: practical and legal considerations for genetics and genomics research in Europe and internationally. *Genome Med* 2017;**9**(1):58.
90. Stark Z, Tan TY, Chong B, *et al*. A prospective evaluation of WES as a first-tier molecular test in infants with suspected monogenic disorders. *Genet Med* 2016;**18**:1090–6.
91. Stark Z, Schofield D, Alam K, *et al*. Prospective comparison of the cost-effectiveness of clinical whole-exome sequencing

with that of usual care overwhelmingly supports early use and reimbursement. *Genet Med* 2017;**19**(8):867–74.
92. Khoury MJ, Iademarco MF, Riley WT. Precision public health for the era of precision medicine. *Ame J Prevent Med* 2016; **50**(3):398–401.
93. Lambin P, Rios-Velazquez E, Leijenaar R, *et al*. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012;**48**(4):441–6.
94. Caudell JJ, Torres-Roca JF, Gillies RJ, *et al*. The future of personalised radiotherapy for head and neck cancer. *Lancet Oncol* 2017;**18**(5):e266–73.
95. Lee G, Lee HY, Park H, *et al*. Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: state of the art. *Eur J Radiol* 2017;**86**: 297–307.