Seminars in Cancer Biology 84 (2022) 16–22

Contents lists available at ScienceDirect

Seminars in Cancer Biology

journal homepage: www.elsevier.com/locate/semcancer

Review

Analytical demands to use whole-genome sequencing in precision oncology

Manja Meggendorfer [a], Vaidehi Jobanputra [b,c], Kazimierz O. Wrzeszczynski [b], Paul Roepman [d], Ewart de Bruijn [d], Edwin Cuppen [d,e], Reinhard Buttner [f], Carlos Caldas [g], Sean Grimmond [h], Charles G. Mullighan [i], Olivier Elemento [j,k], Richard Rosenquist [l,m], Anna Schuh [n], Torsten Haferlach [a,*,1]

[a] MLL Munich Leukemia Laboratory, Munich, Germany
[b] New York Genome Center, 101 Avenue of the Americas, New York, USA
[c] Columbia University Medical Center, 650 W 168th St, New York, USA
[d] Hartwig Medical Foundation, Amsterdam, the Netherlands
[e] Center for Molecular Medicine and Oncode Institute, University Medical Center, Utrecht, the Netherlands
[f] Institute of Pathology, University Hospital Cologne, Germany
[g] Cancer Research UK Cambridge Institute and Department of Oncology, University of Cambridge, United Kingdom
[h] Centre for Cancer Research, University of Melbourne, Melbourne, Australia
[i] Department of Pathology, St. Jude Children's Research Hospital, USA
[j] Institute for Computational Biomedicine, Weill Cornell Medicine, New York, USA
[k] Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, USA
[l] Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden
[m] Department of Clinical Genetics, Karolinska University Hospital, Solna, Sweden
[n] NIHR Oxford Biomedical Research Centre and Department of Oncology, University of Oxford, Oxford, United Kingdom

ARTICLE INFO

Keywords:
Whole-genome sequencing
Clinical WGS
Analytical validation
WGS in routine diagnostics
Precision oncology

ABSTRACT

Interrogating the tumor genome in its entirety by whole-genome sequencing (WGS) offers an unprecedented insight into the biology and pathogenesis of cancer, with potential impact on diagnostics, prognostication and therapy selection. WGS is able to detect sequence as well as structural variants and thereby combines central domains of cytogenetics and molecular genetics. Given the potential of WGS in directing targeted therapeutics and clinical decision-making, we envision a gradual transition of the method from research to clinical routine. This review is one out of three within this issue aimed at facilitating this effort, by discussing in-depth analytical validation, clinical interpretation and clinical utility of WGS. The review highlights the requirements for implementing, validating and maintaining a clinical WGS pipeline to obtain high-quality patient-specific data in accordance with the local regulatory landscape. Every step of the WGS pipeline, which includes DNA extraction, library preparation, sequencing, bioinformatics analysis, and data storage, is considered with respect to its logistics, necessities, potential pitfalls, and the required quality management. WGS is likely to drive clinical diagnostics and patient care forward, if requirements and challenges of the technique are recognized and met.

1. Introduction

Within a decade, next-generation sequencing (NGS) has fundamentally changed the field of oncology, from bench to bedside. Targeted sequencing approaches, i.e., mutation analysis of selected cancer-associated genes, represent the current gold standard in clinical oncology. However, given continuous improvements in NGS throughput, the potential of NGS exceeds targeted approaches and allows the sequencing of entire genomes. Research efforts have illustrated the value of whole-genome sequencing (WGS) [1–11] and its gradual transition into clinical routine will benefit individual patients as well as our understanding of cancer pathogenesis and biology. Due to the complexity of the method and data collected, the implementation, validation, and quality management of the WGS workflow as well as bioinformatics analysis and data storage are challenging. This review aims to identify the analytical demands and to aid the reader in

* Corresponding author.
E-mail address: torsten.haferlach@mll.com (T. Haferlach).
[1] URL: www.mll.com.

https://doi.org/10.1016/j.semcancer.2021.06.009
Received 20 January 2021; Received in revised form 27 May 2021; Accepted 6 June 2021
Available online 10 June 2021

addressing them successfully.

2. Sample availability and material selection

Although single gene testing for single nucleotide variants (SNVs) and insertions and deletions (indels) uses only a few nanograms of DNA input material, current conventional multimodality testing that sequentially tests for an ever-increasing number of single gene abnormalities can rapidly exhaust the low amount of material available from tumor specimens [12]. WGS uses between 50 ng and 1 μg of high-quality DNA (A260/A280: 1.7–2.0) for PCR-amplified or non-amplified libraries and allows interrogation of all different variant types in a single test, as well as future re-analysis and validation of additional test readouts. Here, PCR-free libraries are preferable to avoid amplification bias such as skewed coverage for first exons and GC-rich regions. Stringent quality controls (QCs) have to be applied for DNA concentration, purity, integrity, and fragmentation to produce reliable WGS results (Table 1).

Depending on the sample extraction procedure and tumor type, different materials can be available. However, the most common sample types are formalin-fixed paraffin-embedded (FFPE) and fresh frozen (FF) tissues as well as peripheral blood or bone marrow aspirates. When a biopsy is taken from a patient, the pathologist archives some material as FFPE tissue which can be used for additional analysis at a later time point or for re-analysis. However, FFPE does not represent an ideal source material, as nucleic acids extracted from such samples are fragmented and chemically modified due to age, fixation conditions, and DNA-protein crosslinking resulting in low quality sample libraries which in turn heavily impact downstream analyses. A significant proportion of called variants in FFPE samples are T>C/G>A, which have been reported as artefacts [13]. Formalin fixation is detrimental to the interpretation of SNVs/indels in low complexity regions and prohibits the analysis of copy-number alterations.

Detailed comparison between nucleic acid extraction from frozen and FFPE specimens, library preparation, and high-throughput sequencing including assay validation has been previously described showing that formalin-induced DNA artefacts make WGS interpretation difficult [14–16]. With respect to the obtained sequencing data, comparisons between FF and FFPE specimens have shown that FFPE samples can be interrogated using targeted sequencing including whole-exome approaches [17–23]. A concordance of 70 %–80 % for variants following pairwise analysis of FF and FFPE tissue-derived WES data has been observed, if samples were stored for a maximum of three years [17].

A variety of approaches aim to reduce FFPE-sequencing artefacts. Experimental efforts to this end include improved deparaffinization, repair strategies, and optimized DNA extraction protocols using magnetic bead technology. In order to level the typically low DNA quality, DNA quantity and integrity (= percentage of unfragmented DNA) have to be accurately measured. For a successful library preparation and to prevent excessive loss of material due to size selection steps in the following library preparation, half of the fragments should have a length >120 bp, and one fifth should be >300 bp (e.g. indicated by KAPA Hg

Table 1
Overview of analytical validation. The different steps of test design and quality management, the respective quality assessment and challenges are listed.

	Description	Quality assessment	Comments	Challenges
Test design				
Sample material	FFPE, FF, Blood, Bone marrow	- Tumor burden - Extracted DNA yield - Purity (A260/A280) - Fragmentation	- FF provides better quality - FFPE material should be avoided	- Identification and elimination of material-specific artefacts - Sample availability
Normal control sample	Solid tumors: normal tissue, blood; leukemias: buccal swabs/saliva, sorted T-cells	- Extracted DNA yield - Purity (A260/A280) - Fragmentation	- Necessary to reduce false positive results and to simplify variant interpretation	- Risk of contaminated material for leukemias - Additional sequencing costs
Library	PCR, PCR-free		- PCR Amplification allows low input	- PCR artefacts by amplification
Library preparation	Automation, manual	- Library yield - Fragment length - Homogeneity/uniformity - Reference sample/internal control	- Automation significantly reduces human bias and increases the homogeneity of results - Strict QC is mandatory	- Fast and accurate sample processing - Workflow standardization and reproducibility of results
Sequencing technology	Short read, Long read	- Error rate - Q30 values - Cluster PF - Base composition	- Short reads are cheaper and more accurate but do show limitations to resolve complex loci	- in silico correction of the introduced errors by long read sequencing
Coverage	Low coverage, 30X, >60X		- Sensitivity is mainly influenced by sequencing depth	- Detection of subclonal events
Bioinformatics analysis	Variant calling pipelines	- Demultiplexing efficacy - % aligned - Sequencing depth - Uniformity of coverage	- QC at every step is mandatory to provide high quality data for downstream analyses	- Uniformity of coverage provides a great challenge for reliable variant calling
Bioinformatics infrastructure	pipeline validation/ maintenance	- Electronic reference data file - Software environments: development, testing, production	- Pipelines should constantly be monitored, maintained, and updated if necessary	- Keeping pipelines up-to-date - Data storage - Processing time
Cost	Flexible, fix		- Depend on many factors and should include all steps of the workflow - Sequencing coverage largely influences costs	- Limiting the costs by simultaneously providing high quality results
Turn-around time (TAT)	10−14 days	- Regular review	- TATs are too long for urgent diagnoses	- Reduce sequencing and data analysis time
Quality management				
Regularly reviewed as LDT	Internal audit, external audit	- Test design - Staff training - Equipment maintenance - Standard operating procedure (SOP) - Documentation of controls	- In the US regulated and monitored by CLIA - In the EU regulated and monitored by the individual countries	- Find a secure and viable regulation for data storage, accessibility, and data sharing

DNA QC Kit, Roche Sequencing Systems). For highly degraded DNA samples the DNA input and/or PCR cycles should be increased (DNA input >300 ng, +2 PCR cycles) for the library preparation.

Still, the genomic analysis is more challenging than for non-FFPE derived DNA and requires additional bioinformatic solutions and filtering steps, which may come with a sensitivity and specificity reduction. In addition, FFPE samples are not ideally suited for more complex analysis like fusion transcript, structural variants, loss of heterozygosity analysis, and mutational signature. For high-quality comprehensive characterization, FF tissue is strongly preferred as the input material. Large-scale programs only use FF material as input for WGS [14,24].

For solid tumors, the reliable usability of the available sample material is a crucial prerequisite for clinical implementation of WGS; further optimization efforts are therefore needed. Furthermore, a recent study recommends morphological correlations to address pre-malignant lesions and uncertainty of tumor heterogeneity to reduce errors in tissue sampling and maximize data acquisition [8]. In leukemia diagnostics, sample availability is less of an issue, as peripheral blood and bone marrow are the sample specimens of choice and easily collectable. All the difficulties mentioned above are negligible here, as both the amount of material and the further processing in DNA and RNA are no obstacle and of good quality for downstream analysis like WGS.

3. Sequencing of matched normal samples

For all types of broader gene testing, sequencing matched tumor and normal samples is strongly advised, and perhaps even essential, to optimally differentiate between somatically acquired and tumor variants that are also present in the patient germline. Compared with an assembled consensus reference genome, a typical human genome contains several millions of variants. The contribution of cancer-specific variants is only thousands to hundreds of thousands, depending on tumor type. Incomplete filtering of germline variants, even when this process is >99 % accurate, may result in tens of thousands of false positives, strongly obscuring analysis results or resulting in excessive down-stream interpretation and curation efforts before reporting.

There has been a long-standing discussion regarding the most optimal tissue to use as germline DNA for matched normal sequencing. It's particularly challenging for hematological malignancies due to the high risk of contamination, as leukemic cells may be present in most tissue types. Various solutions have been proposed as sources, including buccal swabs/saliva, nails, hair follicles, fibroblast culturing or direct DNA preparation of skin biopsies, and cell-sorting of peripheral blood T-cells, all of which are associated with specific advantages and shortfalls. Since buccal swaps are easily obtained from the patient, this material is often used as a germline control. However, buccal swaps can be contaminated with blood cells, increasing the risk that relevant somatic mutations might be missed. These false negatives can be prevented by evaluation and comparison of the control's complete mutational profile. Sorted T-cells, on the other hand, are usually very pure, but are only useful as a germline control for myeloid diseases. For solid tumors, tumor cell contamination of the germline sample is usually not an issue and therefore DNA from as less as 200 µL of peripheral blood for example can be used as germline control (Table 1).

Finally, while evaluated WGS data and accompanied statistics are still rare [14,25,26], great efforts are underway to establish tumor-only pipelines, both by the sequencing providing industry and research groups. Such pipelines rely on sophisticated error models and filter strategies and could eliminate the need for germline controls.

4. Library preparation, automation, and reproducibility

NGS is a complex process and one of its key requirements is the library preparation, the quality of which heavily influences the outcome of the sequencing process and usability of the resulting data. NGS technology has been on the market for more than a decade and most molecular diagnostic laboratories already handle library preparation of at least smaller gene panels. However, sampling and library preparation requires a controlled environment and a high level of standardization is necessary to reduce manual bias and increase homogeneity, reproducibility, and efficacy. A key feature of a reliable workflow is integration of quality controls at various steps to monitor sample parameters during the sample preparation to obtain reliable results and increase reproducibility. Therefore, the library yield, the fragment length of the library, the fragment length homogeneity and uniformity are assessed. In 2015 the first reference sample of DNA, the `genome in a bottle` was released, one of the most studied and characterized human sample for single-nucleotide polymorphisms (SNPs) [27]. The periodical preparation of a reference sample (positive control) might be cost intensive, thus other internal controls like the error rate within the highly conserved mitochondrial DNA within a clinical sample are discussed (Table 1). In addition, the set-up of an international External quality assessment (EQA) scheme (an inter-laboratory reproducibility RING-study) is essential to reliably assess the quality of clinical grade WGS and meet ISO standardization.

Automation of a WGS workflow results in more homogenous library concentrations that evenly represent and cover the entire human genome [28]. Various library kits for different platforms are available that are optimized for automation in terms of liquid handling and reagent consumption. The automated systems are sufficiently flexible to adapt to evolving needs (e.g., changes in material), method optimizations, and combination of methods. Automated systems support reagent lot tracking and batch management by automated barcode verifications and logging. The process also guarantees the correct usage of the reagents at the different steps. A key step of WGS library preparations is the addition of unique index sequences (i.e., indexing) per sample that allow multiple libraries to be pooled and sequenced together, with automation ensuring correct index assignment.

Automated workflows and analysis pipelines have to be implemented, tested, and validated to ensure high sensitivity and specificity. Those systems have to be maintained continuously to suitably adapt to necessary optimizations or modifications. Despite the advantages of automation some challenges remain. Due to comparatively high dead volumes and pipetting losses, reagent consumption and accompanied material costs exceed the manual procedure, pipetting volumes are rather small especially for critical enzymatic steps, and the automation system sometimes reaches its limits in these processes.

5. Sequencing technologies

The development of NGS technology has revolutionized our ability to reveal the genomic changes in cancer at unprecedented speed [29]. Large sequencing platforms such as Illumina's HiSeq and NovaSeq series are recommended for up to 250 bp paired-end reads with comparably high sequencing depth. This technology is based on short-read sequencing, hence, significant technical challenges remain, in particular with analysis of loci such as the immunoglobulin or HLA loci that only poorly align to the reference genome and some complex karyotypic aberrations [30,31]. Single molecule sequencing, also called third generation sequencing and currently offered commercially by Pacific BioScience and Oxford Nanopore Technology, permits sequencing of non-amplified native DNA of exceptionally long linear read lengths (1–100 kbp) and fast sequencing times (2–10 h), although requirement for amounts and quality of input material may be prohibitive for specific samples obtained in routine diagnostics. Both technologies have higher (3 %–15 %) and different error rates (mostly indels) than short-read NGS platforms (0.1 %–2.6 % for Illumina instruments) [32]; however, as these errors are relatively random due to the nature of single-molecule sequencing, accurate consensus reads can be achieved bioinformatically provided there is sufficient sequencing depth. At present, long-read sequencing is considerably more expensive than short-read

sequencing; this combined with the high error rates for some variant types means it is currently positioned more as a potentially complementary than alternative approach.

6. Depth-of-coverage and tumor content

Depth-of-coverage and tumor content directly impact the ability to call subclonal variants from WGS data. However, there is a lack of standardization of these parameters. For depth-of-coverage, the first large-scale cancer WES projects, including the International Cancer Genome Consortium (ICGC) [33], used a minimum average coverage of 30X for both tumor and normal samples. However, since then it has been shown that low tumor cell content and purity reduce the sensitivity of calling clonal and subclonal somatic variants [34]. In general, with 10 variant supporting reads and a coverage of 100X, a detection limit of 10 % variant allele frequency (VAF) and therefore 20 % tumor cell content is achieved in best case, which is why in targeted sequencing assays the chosen coverage often exceeds 1000X to reach the technical NGS sensitivity limit of 1 % [35–38].

For WGS, the current view is that the tumor sample should be sequenced to an average depth of ≥90–100X, while the germline sample is sequenced to ≥30X; there are ongoing discussions around whether the germline sample should be sequenced at a higher depth as well. Higher coverage in the tumor than in the matched normal requires additional filtering steps like filtering against normal controls to remove artifacts or heterozygous germline variants that are not covered by any sequencing read in the control sample due to stochastic effects from the somatic variant calls [34]. In cases with low tumor purity, it is also possible to increase the sequencing depth of WGS, but this will unavoidably increase the sequencing cost. Alternatively, the sensitivity of detection of known oncogenic hotspot mutations can be increased by fine-tuning the bioinformatics analysis for such known cancer-related positions as routine somatic variant caller balance between specificity and sensitivity at a genome-wide level.

For tumor content, ICGC required a tumor cell content defined by histopathological examination of at least 60 %; Genomics England allowed 40 %. However, the tumor content determined from histological analysis is often inconsistent and considerably higher than the tumor content determined from sequencing analysis. To date, different large-scale genome studies vary in their requirements for tumor content and coverage (Table 2).

In a clinical setting where tumors with low purity may be examined and even subclonal events need to be detected or ruled out, higher coverage such as >150X per base for the haploid tumor genome may be needed. In practice, macro/micro dissection of solid tumor or cell sorting by flow cytometry/magnetic beads for hematological malignancies before library preparation and sequencing therefore has to be implemented although this further increases costs and turnaround times. A higher sequencing depth and the use of modified bioinformatics methods to detect variants with low mutant allele frequency can compensate for low tumor content, but the challenge of discriminating false-positive variants becomes an issue.

Taking into account all these considerations and as described previously [39], we recommend a minimum of 80–90X mean coverage for the tumor and 30X mean coverage for the matched normal specimen.

Table 2
Depth of coverage, tumor content, and turn-around times for large-scale WGS cancer projects.

	Genomics England 100 000 genomes project	Hartwig Medical Foundation, Hartwig Medical Database	MLL 5000 genome project
No. of genomes	100 000[a]	To date ~ 5200	5000
Samples analyzed	Rare diseases and tumors	Metastatic tumors	Leukemia & Lymphoma
Tumor content (minimum)	40 %	20 %	20 %
Average coverage	80X	90X	100X
Turnaround time	21 days	9–16 days	8–9 days

[a] 50 000 genomes from 25 000 cancer patients, with two genomes per patient (tumor vs. matched normal), and 50 000 genomes from ~17 000 rare disease patients and their relatives (patient plus two blood relatives). Accessed via: https://www.genomicsengland.co.uk/about-genomics-england/the-100000-genomes-project/, accessed on April 24th 2020.

7. Bioinformatic analysis

In a clinical (hospital) setting, the ICT infrastructure and bioinformatics skill/availability are too often insufficient. A bioinformatics analysis comprises the different data processing steps from the raw sequencing reads (FASTQ files) to the variant report files (annotated VCF files) generated for clinical interpretation.

No single algorithm or software program is sufficient to perform all the different steps of the workflow and, hence, a linear combination of informatics tools is needed for comprehensive data analysis (i.e., bioinformatics pipeline). Here, the reads are aligned to the reference genome (excluding reads with low quality scores), sorted, and, following further quality assessment (i.e. % Aligned reads, insert sizes, uniformity of coverage), PCR duplicates are marked or removed before variants are called (Table 1). In addition to SNVs and indels, WGS data also allows the calling of copy number variations, structural variants, microsatellite (in)stability, tumor mutational burden, and other mutational signature (detailed in the Clinical Interpretation paper). However, variant calling algorithms are only optimized to call one type of variant with sufficiently high sensitivity and specificity. So far, no gold standard pipeline exists that is ready-to-use for clinical genomics workflows.

Before and after processing the data, stringent QCs have to be applied to guarantee high-quality accurate results. QC metrics should include as a minimum constant signal intensity, quality scores for base calling and alignment, and the depth and uniformity of coverage for the whole genome. There are many regions in the genome that cannot always be accurately determined by current analysis tools, including regions of high homology for which reads cannot be uniquely mapped, regions where the reference genome contains errors, regions with multiple reference haplotypes, and tandem repeats that extend beyond the sequenced read length. Those problematic regions have to be considered carefully, especially if they overlap with clinically relevant areas, which could lead to false-negative results. Testing and validation of clinical analysis pipelines is non-negotiable to ensure accurate and reproducible results. During the implementation and validation of the pipeline the laboratories should account for systematic errors of the sequencing platform, the aligner, and the variant callers. For pipeline validation, a set of characterized sequence variations, including disease-associated sequence variations, is needed to evaluate the pipelines capability to reliably call those variants without generating false-positive results. Furthermore, the `genome in a bottle` reference set of variant calls might work as control for SNP calling [27]. It is desirable to establish an electronic reference data file collection as a gold standard for re-evaluations. The pipeline has to be maintained continuously to account for periodical updates of the reference genome, the aligner, and the variant callers. This usually requires several software environments, e.g., one for development and testing and one for production. Each software change has to be documented and validated to ensure accurate variant calling.

Multiple repositories for genomic data exist to deposit, search, analyze and download cancer data, all following the FAIR (Findable, Accessible, Interoperable, Reusable) principle (e.g. Gene expression omnibus: https://www.ncbi.nlm.nih.gov/geo/, European Genome-phenome Archive: https://ega-archive.org/, Genomic Data Commons: https://portal.gdc.cancer.gov/, GWAS Catalog: https://www.ebi.ac.uk/

gwas/) although only a handful of studies and datasets are publicly available so far. The collection and harmonization of WGS data sets would greatly benefit unified variant annotation and would facilitate the generation of standardized pipelines and processing guidelines, providing instructions to less experienced analysts and smaller institutions.

8. Bioinformatics infrastructure and data storage

WGS data have higher demands by several orders of magnitude on the bioinformatics analysis compared to WES or targeted NGS data due to heavier computational load, increased number of variants for analysis, and larger data archives. Regarding costs for bioinformatics and storage, short-term capacities for direct computing power after sequencing (from FASTQ to VCF) and long-term storage capacities must be created. The developed infrastructure has to be able to process and evaluate the data in a reasonable time frame and should guarantee secure long-time data storage and the centralization of results to simplify quality checks and variant annotation. Such infrastructure requires the installation of a high-performing computing cluster, possibly filling an entire data center with emphasis on flexible but controlled access and data security. Hence, based on the gained experience of large-scale sequencing projects, various countries (e.g. UK, Sweden, Estonia) have implemented infrastructures that allow for a centralized and harmonized data processing and interpretation, shifting the data analysis to specialized centers. To save on the space and costs that come with the development and maintenance of such an IT-infrastructure, cloud computing should be considered. Especially hospitals and institutions located in countries lacking an elaborate health data infrastructure could greatly benefit from cloud computing by not only renting computational power and data storage capacity in a cost-effective way but also by borrowing bioinformatics expertise through the usage of precompiled pipelines. Medical data security and protection of a patients' privacy are essential in a diagnostic setting also from an ethico-legal perspective. Besides virtually unlimited resources, cloud computing provides the added advantage of up-to-date data security with the Health Insurance Portability and Accountability Act compliance as a holistic solution. Nevertheless, further improvement in data security and privacy are needed for a broad adoption of cloud computing in the healthcare industry [40]. Irrespective of the computing platform selected, emphasis on data security is critical especially if genomic data is to be stored together with other clinical data.

9. Costs

Considerations on costs should include all steps of the WGS protocol, which comprise DNA extraction, library preparation and sequencing, bioinformatics, clinical interpretation/reporting, and long-term storage of raw and processed data, both with respect to the necessary equipment and highly trained personnel. Currently, the sequencing cost of a 30X genome is estimated at between $500–$2500 worldwide. However, the cost for providing a clinical diagnosis from WGS is much higher depending on turnaround time and level of interpretation. For somatic mutation analysis the sequencing cost of the genome has to be doubled to account for tumor-normal pairs. In the oncology space, only one study [41] so far performed a comprehensive micro-costing exercise of the real cost of WGS for single centers and established an average cost of over £6000 per cancer genome.

10. Turnaround times

When considering turnaround time, it is important to note the necessary balance between a fast and a comprehensive diagnosis. There are diseases where a rapid diagnosis is potentially life-saving. One example is acute promyelocytic leukemia whose diagnosis has to be established within hours and directs therapy with all-trans retinoic acid (ATRA). The UK's NHS have adopted a two-phase approach that allows urgent testing for the pathognomonic *PML-RARA* translocations by a targeted single gene method. WGS is unlikely to assist diagnostics in such cases. However, it can provide valuable insight into tumors that are known for their genetic complexity and lead to the identification of pathogenic lesions, which, with the advance of precision medicine, might be suitable therapeutic targets. There are first examples on the coordination of findings from genetic characterization with therapeutic decisions (detailed in the Clinical Utility paper).

For solid tumors, most patients are not eligible for alternatives to the standard-of-care approach until treatment with that standard fails and they have relapsed disease. In patients with a biopsy sample of a relapsed tumor available, this is an optimal assay sample for identifying a specific alternative treatment approach, but more rapid turnaround is required for this information to be utilized clinically [42]. However, routinely obtaining a biopsy sample from a patient presenting with metastatic cancer can be difficult. Thus, having data from the primary diagnosis can help determine new options for the patient and aid therapeutic decisions in a timely manner.

Time required to thoroughly review and interpret the different variant types detected by this comprehensive testing and describing the implicated therapeutic, diagnostic, or prognostic implications in the reports increases the effort required. Therefore, there is variation among the laboratories returning the information in a clinically relevant time frame. The same holds true for large-scale WGS cancer studies (Table 2).

In the clinical routine, most patients and physicians would find 10–14 days as an acceptable time for results that might inform treatment decisions. Some centers may perform rapid turnaround time versus others who investigate the primary diagnostic biopsy sample or resection sample in a lenient time frame, in which any treatment-relevant information is included in the patient's medical record and then taken into consideration if and/or when the patient develops a recurrent or metastatic cancer. In some centers, a 2-step approach may therefore reasonable, a primary rapid tumor only analysis followed by an all-encompassing tumor/normal analysis. In this way, diagnostic or therapy-relevant genetic alterations could be detected very quickly in a targeted manner, followed by a more detailed and comprehensive but time-consuming characterization.

11. Quality management

In the United States, diagnostic tests provided to clinical laboratories are regulated by the US Food and Drug Administration (FDA). Although, WGS in a clinical setting for human specimens is not currently FDA-approved, in-house laboratory developed tests (LDTs), which are regulated by the Clinical Laboratory Improvement Amendments (CLIA), are already in use, e.g., St Jude's Children Hospital (CLIA certified). While in Europe the coupling of companion diagnostic test to a specific drug is no that strictly enforced by the European Medicine Agency (EMA) like in the US, LDTs are much more frequently used instead of companion diagnostic tests, e.g. the UK's Genomics England (ISO 15189 accredited, UKAS certified), and the Hartwig Medical Foundation in The Netherlands (ISO 17025 accredited) have WGS pipelines approved as LDTs. Therefore the quality assurance schemes for LDTs are even more important. Laboratories offering LDTs are subject to specific laboratory standards governing certification, personnel, proficiency testing, patient test management, quality assurance, QC and inspections to establish the analytical validity of the developed test. This includes accuracy, precision, analytical sensitivity and specificity, reportable range, reference range or intervals, and other performance-relevant metrics. Laboratories may also comply with quality standards described in the international standard ISO 15189 "Medical laboratories - Requirements for quality and competence", which is based on ISO/IEC 17025 and ISO 9001, and contains special requirements for the quality and competence of medical laboratories.

A new European In Vitro Diagnostics Regulation (IVDR) was

officially published and came into force on 26th May 2017. A transition period of five years applies to manufacturers of already approved medical devices to meet the IVDR requirements. The new IVDR differs from the past EU IVD Directive in several important aspects. The main changes include extension of the scope, including high-risk products manufactured and used in a single healthcare facility, IVD for diagnostic (including internet-based) purposes, genetic testing and other tests to provide information on the predisposition of patients to a particular disease or the effect of treatment, and stricter requirements for technical documentation and clinical evaluation [43]. While the practical implementation of the new guidelines is still being discussed, the individual countries are currently responsible for quality assurance. While the analytical parts, the technical generation of molecular genetic data by guidelines and recommendations are well specified, data collection and data storage are not yet included. The latter have yet to be implemented as best practice guidelines (Table 1).

12. Conclusions

While the method of WGS is demanding from its implementation to the analysis and storage of WGS data, its potential outweighs the analytical challenges. Research initiatives continue to prove the value of WGS, which starts with improving individual patient care and extends to the advancement of classification, prognosis, and therapy. In consideration of declining sequencing costs and the increasing importance of precision diagnostics and targeted therapeutics in clinical oncology, we envision the transition of WGS from research to routine. This transition should be accompanied by a comprehensive and critical assessment of the challenges and advantages of the method; the reader is therefore also referred to the in-depth discussion on its Clinical Utility and Clinical Interpretation in this review series.

Funding source

The research was funded/supported by the National Institute for Health Research (NIHR)Oxford Biomedical Research Centre (BRC). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. RR received funding from the Swedish Cancer Society, the Swedish Research Council, the Knut and Alice Wallenberg Foundation, Karolinska Institutet, Karolinska University Hospital, and Radiumhemmets Forsknings fonder.

Declaration of Competing Interest

TH is part owner of MLL Munich Leukemia Laboratory. MM is employed by MLL. AS: honoraria from Gilead, Roche, Janssen, Abbvie, Astra Zeneca, Adaptive Biotechnology, Jazz, Base Genomics, Illumina, Oxford Nanopore Technology. Unrestricted educational grants from Astra Zeneca and Janssen. In-kind contributions from Illumina and Oxford Nanopore Technology. RB: Honoraria from Abbvie, Amgen, Astra Zeneca, BMS, Illumina, MSD, Merck-Serono, Lilly, Roche. RR: honoraria from AbbVie, AstraZeneca, Illumina, Janssen, and Roche. CGM: honoraria from Amgen, Illumina; research funding from Pfizer, Abbvie. CC is a member of the External Science Panel of AstraZeneca iMED, honoraria from Illumina and research grants (administered by the University of Cambridge) from AstraZeneca, Sevier, Genentech and Roche. OE is co-founder and equity holder of Volastra Therapeutics and One Three Biotech, serves on the scientific advisory board and holds equity in Freenome and Owkin and receives research funding from Janssen and Eli Lilly.

Acknowledgements

The authors would like to thank Kristine Jinnett (Illumina, Inc.) for her assistance in the preparation of the manuscript.

References

[1] A. Schuh, J. Becq, S. Humphray, et al., Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns, Blood 120 (November (20)) (2012) 4191–4196, https://doi.org/10.1182/blood-2012-05-433540.
[2] L. Ding, T.J. Ley, D.E. Larson, et al., Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing, Nature 481 (January (7382)) (2012) 506–510, https://doi.org/10.1038/nature10738.
[3] E.D. Pleasance, R.K. Cheetham, P.J. Stephens, et al., A comprehensive catalogue of somatic mutations from a human cancer genome, Nature 463 (January (7278)) (2010) 191–196, https://doi.org/10.1038/nature08658.
[4] M.J. Ellis, L. Ding, D. Shen, et al., Whole-genome analysis informs breast cancer response to aromatase inhibition, Nature 486 (June (7403)) (2012) 353–360, https://doi.org/10.1038/nature11143.
[5] Pan-cancer analysis of whole genomes, Nature 578 (February (7793)) (2020) 82–93, https://doi.org/10.1038/s41586-020-1969-6.
[6] S. Nik-Zainal, H. Davies, J. Staaf, et al., Landscape of somatic mutations in 560 breast cancer whole-genome sequences, Nature 534 (June (7605)) (2016) 47–54, https://doi.org/10.1038/nature17676.
[7] A. Schuh, H. Dreau, S.J.L. Knight, et al., Clinically actionable mutation profiles in patients with cancer identified by whole-genome sequencing, Cold Spring Harb. Mol. Case Stud. 4 (April (2)) (2018), https://doi.org/10.1101/mcs.a002279.
[8] G. Echejoh, Y. Liu, G. Chung-Faye, et al., Validity of whole genomes sequencing results in neoplasms in precision medicine, J. Clin. Pathol. 29 (October) (2020), https://doi.org/10.1136/jclinpath-2020-206998.
[9] Q. Ma, J. Wang, J. Qi, et al., Increased chromosomal instability characterizes metastatic renal cell carcinoma, Transl. Oncol. 14 (November (1)) (2020), 100929, https://doi.org/10.1016/j.tranon.2020.100929.
[10] M. Wong, C. Mayoh, L.M.S. Lau, et al., Whole genome, transcriptome and methylome profiling enhances actionable target discovery in high-risk pediatric cancer, Nat. Med. 26 (November (11)) (2020) 1742–1753, https://doi.org/10.1038/s41591-020-1072-4.
[11] M.K. Samur, A. Aktas Samur, M. Fulciniti, et al., Genome-wide somatic alterations in multiple myeloma reveal a superior outcome group, J. Clin. Oncol. 38 (September (27)) (2020) 3107–3118, https://doi.org/10.1200/jco.20.00461.
[12] M. Kerick, M. Isau, B. Timmermann, et al., Targeted high throughput sequencing in clinical cancer settings: formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity, BMC Med. Genomics 4 (September (29)) (2011) 68, https://doi.org/10.1186/1755-8794-4-68.
[13] A. Marchetti, L. Felicioni, F. Buttitta, Assessing EGFR mutations, N. Engl. J. Med. 354 (February (5)) (2006) 526–528, https://doi.org/10.1056/NEJMc052564, author reply 526-528.
[14] P. Robbe, N. Popitsch, S.J.L. Knight, et al., Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 genomes project, Genet. Med. 20 (October (10)) (2018) 1196–1205, https://doi.org/10.1038/gim.2017.241.
[15] H. Do, A. Dobrovic, Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization, Clin. Chem. 61 (January (1)) (2015) 64–71, https://doi.org/10.1373/clinchem.2014.223040.
[16] S.Q. Wong, J. Li, A.Y. Tan, et al., Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing, BMC Med. Genomics 7 (May (13)) (2014) 23, https://doi.org/10.1186/1755-8794-7-23.
[17] J. Hedegaard, K. Thorsen, M.K. Lund, et al., Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue, PLoS One 9 (5) (2014), e98187, https://doi.org/10.1371/journal.pone.0098187.
[18] E.M. Van Allen, N. Wagle, P. Stojanov, et al., Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine, Nat. Med. 20 (June (6)) (2014) 682–688, https://doi.org/10.1038/nm.3559.
[19] S. Munchel, Y. Hoang, Y. Zhao, et al., Targeted or whole genome sequencing of formalin fixed tissue samples: potential applications in cancer genomics, Oncotarget 6 (September (28)) (2015) 25943–25961, https://doi.org/10.18632/oncotarget.4671.
[20] A. Astolfi, M. Urbini, V. Indio, et al., Whole exome sequencing (WES) on formalin-fixed, paraffin-embedded (FFPE) tumor tissue in gastrointestinal stromal tumors (GIST), BMC Genomics 16 (November (3)) (2015) 892, https://doi.org/10.1186/s12864-015-1982-6.
[21] E. Oh, Y.L. Choi, M.J. Kwon, et al., Comparison of accuracy of whole-exome sequencing with formalin-fixed paraffin-embedded and fresh frozen tissue samples, PLoS One 10 (12) (2015), e0144162, https://doi.org/10.1371/journal.pone.0144162.
[22] R. De Paoli-Iseppi, P.A. Johansson, A.M. Menzies, et al., Comparison of whole-exome sequencing of matched fresh and formalin fixed paraffin embedded melanoma tumours: implications for clinical decision making, Pathology 48 (April (3)) (2016) 261–266, https://doi.org/10.1016/j.pathol.2016.01.001.
[23] H. Rennert, K. Eng, T. Zhang, et al., Development and validation of a whole-exome sequencing test for simultaneous detection of point mutations, indels and copy-number alterations for precision cancer care, NPJ Genom. Med. 1 (2016), 16019, https://doi.org/10.1038/npjgenmed.2016.19.
[24] P. Priestley, J. Baber, M.P. Lolkema, et al., Pan-cancer whole-genome analyses of metastatic solid tumours, Nature 575 (November (7781)) (2019) 210–216, https://doi.org/10.1038/s41586-019-1689-y.

[25] H.M. Wood, O. Belvedere, C. Conway, et al., Using next-generation sequencing for high resolution multiplex analysis of copy number variation from nanogram quantities of DNA from formalin-fixed paraffin-embedded specimens, Nucleic Acids Res. 38 (August (14)) (2010) e151, https://doi.org/10.1093/nar/gkq510.
[26] M.R. Schweiger, M. Kerick, B. Timmermann, et al., Genome-wide massively parallel sequencing of formaldehyde fixed-paraffin embedded (FFPE) tumor tissues for copy-number- and mutation-analysis, PLoS One 4 (5) (2009) e5548, https://doi.org/10.1371/journal.pone.0005548.
[27] J.M. Zook, B. Chapman, J. Wang, et al., Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls, Nat. Biotechnol. 32 (March (3)) (2014) 246–251, https://doi.org/10.1038/nbt.2835.
[28] Hamilton Company, Application Note, Lit. No. AN-1804-01 – 04/2018, 2018.
[29] D.R. Bentley, S. Balasubramanian, H.P. Swerdlow, et al., Accurate whole human genome sequencing using reversible terminator chemistry, Nature 456 (November (7218)) (2008) 53–59, https://doi.org/10.1038/nature07517.
[30] A. Ameur, W.P. Kloosterman, M.S. Hestand, Single-molecule sequencing: towards clinical applications, Trends Biotechnol. 37 (January (1)) (2019) 72–85, https://doi.org/10.1016/j.tibtech.2018.07.013.
[31] E.L. van Dijk, Y. Jaszczyszyn, D. Naquin, C. Thermes, The third revolution in sequencing technology, Trends Genet. 34 (September (9)) (2018) 666–681, https://doi.org/10.1016/j.tig.2018.05.008.
[32] F. Pfeiffer, C. Gröber, M. Blank, et al., Systematic evaluation of error rates and causes in short samples in next-generation sequencing, Sci. Rep. 8 (July (1)) (2018), 10950, https://doi.org/10.1038/s41598-018-29325-6.
[33] T.J. Hudson, W. Anderson, A. Artez, et al., International network of cancer genome projects, Nature 464 (April (7291)) (2010) 993–998, https://doi.org/10.1038/nature08987.
[34] T.S. Alioto, I. Buchhalter, S. Derdak, et al., A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing, Nat. Commun. 6 (December (9)) (2015) 10001, https://doi.org/10.1038/ncomms10001.
[35] L.J. Jennings, M.E. Arcila, C. Corless, et al., Guidelines for validation of next-generation sequencing-based oncology panels: a joint consensus recommendation of the association for molecular pathology and college of American pathologists, J. Mol. Diagn. 19 (May (3)) (2017) 341–365, https://doi.org/10.1016/j.jmoldx.2017.01.011.
[36] J.J. Salk, M.W. Schmitt, L.A. Loeb, Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations, Nat. Rev. Genet. 19 (May (5)) (2018) 269–285, https://doi.org/10.1038/nrg.2017.117.
[37] A. Petrackova, M. Vasinek, L. Sedlarikova, et al., Standardization of sequencing coverage depth in NGS: recommendation for detection of clonal and subclonal mutations in cancer diagnostics, Front. Oncol. 9 (2019) 851, https://doi.org/10.3389/fonc.2019.00851.
[38] M.A. Quail, M. Smith, P. Coupland, et al., A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers, BMC Genomics 13 (July (24)) (2012) 341, https://doi.org/10.1186/1471-2164-13-341.
[39] K.O. Wrzeszczynski, V. Felice, A. Abhyankar, et al., Analytical validation of clinical whole-genome and transcriptome sequencing of patient-derived tumors for reporting targetable variants in cancer, J. Mol. Diagn. 20 (November (6)) (2018) 822–835, https://doi.org/10.1016/j.jmoldx.2018.06.007.
[40] Y. Al-Issa, M.A. Ottom, A. Tamrawi, eHealth cloud security challenges: a survey, J. Healthc. Eng. 2019 (2019), 7516035, https://doi.org/10.1155/2019/7516035.
[41] K. Schwarze, J. Buchanan, J.M. Fermont, et al., The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom, Genet. Med. 22 (January (1)) (2020) 85–94, https://doi.org/10.1038/s41436-019-0618-7.
[42] M.F. Berger, E.R. Mardis, The emerging clinical relevance of genomics in cancer medicine, Nat. Rev. Clin. Oncol. 15 (June (6)) (2018) 353–365, https://doi.org/10.1038/s41571-018-0002-6.
[43] G. Dagher, K.F. Becker, S. Bonin, et al., Pre-analytical processes in medical diagnostics: new regulatory requirements and standards, N. Biotechnol. 52 (September (25)) (2019) 121–125, https://doi.org/10.1016/j.nbt.2019.05.002.