Accepted Article

Next Generation Sequencing: A Change of Paradigm in Molecular Diagnostic Validation

Manuel Salto-Tellez[1] and David Gonzalez de Castro[2]

[1]Northern Ireland Molecular Pathology Laboratory, Centre for Cancer Research and Cell Biology, Queen's University Belfast, United Kingdom

[2]Molecular Diagnostics Department, The Centre for Molecular Pathology, The Royal Marsden NHS Foundation Trust, London, UK

Corresponding author:

Prof Manuel Salto-Tellez

Centre for Cancer Research and Cell Biology

Queen's University Belfast

97 Lisburn Road

Belfast BT9 7BL

Telephone: +442890972243

Email: m.salto-tellez@qub.ac.uk

Commentary in relation to Tothill RW, Li J, Mileshkin L, *et al.* Massively-parallel sequencing assists the diagnosis and guided treatment of cancers of unknown primary. . *J Pathol.* 2013 DOI: 10.1002/path.4251.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/path.4365

This article is protected by copyright. All rights reserved

Accepted Article

Abstract

Next Generation Sequencing (NGS) is beginning to show its full potential for diagnostic and therapeutic applications. In particular, it is enunciating its capacity to contribute to a molecular taxonomy of cancer, to be used as a standard approach for diagnostic mutation detection, and to open new treatment options that are not exclusively organ-specific. If this is the case, how much validation is necessary and what should be the validation strategy, when bringing NGS into the diagnostic/clinical practice? This validation strategy should address key issues such as: what is the overall extent of the validation?; should essential indicators of test performance such as sensitivity of specificity be calculated for every target or sample type?; should bioinformatic interpretation approaches be validated with the same rigour?; what is a competitive clinical turnaround time for an NGS-based test and when does it become a cost-effective testing proposition? While we address these and other related topics in this commentary, we also suggest that a single set of international guidelines for the validation and use of NGS technology in routine diagnostics may allow us all to make a much more effective use of resources.

Key words: NGS, validation, technology

The recent paper by Tothill et al. in one of the recent issues of *Journal of Pathology* [1] illustrates an intelligent application of complex genomic information in a very specific clinical problem. Both fresh-frozen and formalin-fixed paraffin embedded samples of patients with cancers of unknown primary (CUP) were analysed with next generation sequencing (NGS) technology. In 75% of the patients tested, the results revealed new therapeutic options, as well as certain signatures that are “etiological” in nature and, as such, are indicative of a

This article is protected by copyright. All rights reserved

Accepted Article

likely tissue/organ of origin. Based on this and other prior studies, the authors suggest a pathway to integrate high-throughput genomic information in the overall therapeutic decision-making. This type of studies are taking us to a new scenario, already predicted when these technologies were made available in their first instance, beyond their clear role in the discovery of new molecular mechanisms of disease. Indeed, the work by Tothill et al, as well as other contemporary papers, is beginning to illustrate the diagnostic and therapeutic application of massive parallel sequencing (MPS) approaches for the reclassification of diagnosis [2], the therapeutic decision-making beyond the classic organ-specific treatment options [1], or the detection of standard-of-care mutations in comparison with the single-gene analysis gold-standard [3].

As readers, peer-reviewers and users of this technology in a diagnostic framework, a question is always at the back of our minds when confronted with these works: how much validation is necessary and what should be the validation strategy, when bringing NGS into the diagnostic/clinical practice? How do we resolve the dichotomy between the clear opportunities that are opened by this technology and the requirements to fully validate it in the routine diagnostic setting? Addressing the necessary criteria for this validation, summarized in Table 1, will dictate the speed and availability of this technology to the wider molecular diagnostic community.

Should every variant be validated?

This is one of the areas where the transition from single gene or low-throughput testing into high-throughput diagnostics is, on paper, more controversial. Should every single variant within the DNA portion of interest be validated? This is simply not possible by design: even with a relatively small NGS panel of 20 genes, the number of possible variants detectable by

This article is protected by copyright. All rights reserved

Accepted Article

NGS can be in the thousands. There are no bio-resources with such large collection of annotated (and validated) variants and the cost would be prohibitive. A more pragmatic approach, as described by several recent studies may be to validate the technology, taking the main clinically relevant targets as a model [4, 5]. The main advantage is that this is an affordable and sustainable solution even if new variants within the region analysed are described in new studies. The main disadvantage is the possible suboptimal detection of large deletions/insertions (i.e. >20bp) in some areas, or the unfaithful representation of certain DNA regions (e.g. GC-rich) leading to suboptimal coverage of certain potential variants.

Should we calculate sensitivity / specificity / precision / limit of detection (LOD) for every target? And for every sample type?

Here we face several challenges, namely cost again but also the intrinsic higher sensitivity of NGS technology compared to Sanger sequencing, the wider range of variants detected in a single test and the inherent artefacts caused by formalin fixation in FFPE-derived DNA. A small NGS amplicon design covering just 20kb of genomic DNA for the detection of key oncogenic mutational hotspots in 10 to 20 genes would be, by design, mutation-rich in the appropriate clinical samples, and therefore can lead to an average of 3-6 non-polymorphic variants per case. For single mutation assays, aiming for a 99% statistical sensitivity (with a 95% confidence interval) would require testing at least 300 specimens for the mutation or region analyzed which, as explained before, is not feasible in terms of cost and access to such samples (for more details regarding a framework validation of genetic tests please refer to Mattocks et al) [6]. In the case of the 20kb NGS panel it may be statistically justified to analyzed only 100 specimens or even less as long as they yield a total of >300 variants that are correctly classified and validated by orthogonal methods to achieve an overall 99% sensitivity for the panel test. Indeed, one could argue that detection of the polymorphic

This article is protected by copyright. All rights reserved

Accepted Article

variants that such panel may contain (10-20 SNP per sample) can help reducing the total number of specimens required for the validation, and this approach has been used for validation of NGS panels using cell lines from 1,000 genomes, HapMap or other well-annotated cohorts. This would in fact be the case for validation of an NGS panel aimed at detection of germline variants, and to assess some of the performance characteristics of the design (e.g. minimum number of reads, percentage of on-target reads, theoretical LOD). However, this approach is not sufficient when applied to detection of somatic mutations (including very low allele frequencies) in FFPE tissue from solid tumour samples. In this latter case, an additional validation step in clinical specimens is needed to assess the performance of the test in an end-to-end setting, and should include samples from all the applicable tumour types at different levels of neoplastic cell content and different levels of cellularity and fragmentation, covering a wide range of clinically suitable specimens.

In most ongoing NGS validations, one can detect –in parallel to those clinically validated mutations that are known to be present in the validation sets– other low-frequency variants of the same genes at much lower percentage, or the common actionable mutations in a small clone previously undetected. How can these low-frequency variants be validated? From a strict analytical viewpoint, the gold-standard platform to validate low-frequency mutations would need to exercise a higher degree of sensitivity than the current gold-standards utilized to date, and is likely to be closer to differential display PCR (ddPCR)-like methods rather than the conventional Sanger sequencing approaches, or a combination of Sanger sequencing and other more sensitive methods such as RQ-PCR.

In the event that we are convinced of their true nature, should they be reported and, if so, what is their degree of clinical relevance? The answer is not easy and would depend very much on the appetite of oncologists to take into account such information, either for clinical

This article is protected by copyright. All rights reserved

Accepted Article

utilization or as an ongoing training exercise in the utilization of complex genomic information in patient care. Thus, a prior agreement between the oncologists and the molecular diagnosticians in each institution on the extent of reporting of genomic abnormalities, as well as the precise wording of these findings within the diagnostic report, may be necessary to make the most of molecular medicine.

Once the validation is in place to apply a NGS test design to a specific cancer type, can we presume that the same analytical validation applies to all cancer types? Again, the Sanger sequencing validation paradigms may come handy. Here we accept that samples of a similar size/volume and a similar pre-analytical treatment may only need small, confirmatory extra verification. For instance, if a *KRAS* mutation testing design works for FFPE tissue from colorectal cancer biopsies, only a verification design (rather than formal validation) of a few lung cancer samples would be necessary to confirm that this test is also applicable to pulmonary adenocarcinoma, which is subjected to a similar pre-analytical protocol.

How do we validate the bioinformatics pipelines and curation of the results?

Does this validation need to be mutation- or panel-specific? Do we need standard bioinformatics packages for diagnostic interpretation? And, if we change the NGS target design, or we change the bioinformatics pipelines, do we need to validate the whole approach all over again?

The necessity to have a high degree of stringency in the bioinformatics approach to diagnostics is very much highlighted by the ISO15189 requirements and accreditation standards for medical laboratories, calling for the same need for validation, maintenance, document control and quality assurance for “dry-lab” interpretative aspects of diagnostics as

This article is protected by copyright. All rights reserved

Accepted Article

for "wet-lab" protocols. Any new pipeline requiring bioinformatics needs validation, and any transfer of existing bioinformatic pipelines needs verification. As such, *diagnostic bioinformatics* is likely to become one of the fastest growing areas in modern medicine. In practical terms, this could be potentially feasible without significant increase in cost, as long as the NGS panel design and protocol has not changed, as there will be raw data (FASTQ files) in the laboratories from the original validations that can be re-analysed dividing the cohort into a "training set" followed by a "validation set". Any new modifications to the analysis pipeline, like with any software developments, needs to adhere to the principles of document control and version release protocols. These principles are expected to be applied in commercially-available analysis packages, in which case, a process of verification as explained before needs to be followed before clinical implementation by the testing laboratory. However, it is likely that many NGS users would develop their own analytical pipelines, to suit the particular needs or protocols of the laboratory, using freely available third-party algorithms and databases that are constantly evolving. In this scenario, a validated analytical pipeline should not be modified or upgraded without the required verification and/or validation steps, including documentation and version control.

Should laboratories decide to go ahead with commercially available panel designs for diagnostic purposes, and in the absence of an FDA or CE-marked product for NGS cancer mutation analysis (at the time in which this commentary is written), there are key aspects that would need to be considered by the user: robustness in FFPE analysis, adequacy of the choice of the commercial panel for the clinical interest of the centre and for its volume of testing, and cost/turnaround time. In any case, even if the provider has extensively validated the commercial panel, the diagnostic laboratory must perform and document a verification procedure in house. If the model pursued is that of in-house NGS testing with an in-house

This article is protected by copyright. All rights reserved

Accepted Article

analysis of the results by a third-party algorithm, or if commercial, extramural bioinformatics interpretation by an expert company is preferred, there are two aspects that would be paramount to make a decision, namely a) the comprehensive, accurate and easy-to-use nature of the third party bioinformatics resource; and b) the level of expertise and the accreditation / certification status in the diagnostic setting that the companies engaged in interpretation of in-house results can document.

An additional consequence of improved analysis pipelines is the potential for new results coming to light in specimens previously tested with the same NGS panel and analysed with a prior version of the software. Diagnostic laboratories will have to define specific policies and criteria for when (and if) the new analysis should be performed in a previously reported case, for example, if the patient progresses under current therapy without a new biopsy being available for analysis, or a change in treatment is being considered on the basis of toxicity.

Is the turnaround time (TAT) competitive?

Table 2 indicates the basic steps for targeted NGS mutation analysis based on amplicon or capture approaches. How does it compare with our current single gene, standard-of-care testing? A single gene test for a clinical case can have a TAT of 2-3 days. However, internal laboratory organizations, to increase economy of scales and throughput (i.e. sample batching), make a realistic TAT closer to 5-7 working days. Hence, in principle the TAT would be consistent with the length of time required for NGS testing -assuming there is only one NGS workflow in the laboratory for all the different tumour types-, which can be as short as 4 working days (see table 2). This is so provided that the last bioinformatics step is fully integrated in this process. Again, it would appear that the need for "diagnostic bioinformaticians" or robust, automated commercially-available analytical solutions

This article is protected by copyright. All rights reserved

Accepted Article

embedded in the fabric of molecular diagnostic laboratories may be a *condition sine qua non* for a competitive NGS diagnostic service. Current guidelines for diagnostic detection of some of the standard-of-care single mutations [7] indicate that "Patients with stage IV lung cancer have median untreated life expectancy of approximately 16 weeks; 20% of this time should not be spent waiting for test results" and that "While this technology has great scientific promise and platform consolidation is logistically appealing, the ability to generate large amounts of data of unproven significance should not take precedence over the timely generation of clinically useful data."

Most of the current uses of NGS in the diagnostic setting are based on the direct amplicon detection approach stated above, suitable for the detection of point mutations and small indels. However, the technology is versatile enough to detect and quantitate RNA (RNA-Seq), as well as other DNA changes such as deletions, insertions or translocations. For the latter, a targeted capture strategy (enriching for a specific sequence prior to NGS analysis by hybridization steps) is currently favored [4, 5], which opens new detection possibilities with the technology but could double the TAT up to 10 working days.

What is the minimum number of targets to make NGS cost-efficient in diagnostics? At the time of writing this commentary, the jury is still out in most healthcare systems as to the extent to which we should test the cancer genome on a regular basis for diagnostic purposes. While the technology theoretically would allows to detect any possible mutation, and oncologists (particularly in the academic environment) would like to see mutation detection profiles that include all possible recurrent changes that are therapeutically actionable, the number of personalised medicine examples per cancer type that are currently reimbursed is still modest. Indeed, some of the recent studies aiming to apply precision medicine across the

This article is protected by copyright. All rights reserved

Accepted Article

board represent a strong reality check: only 13% of breast cancer patients suitable for analysis with both array CGH and Sanger sequencing were suitable for targeted therapies and only 30% of these achieve disease control [8].

A full characterization of all the costs associated with NGS technology in diagnostics is not within the scope of this commentary. However, an analysis of consumable costs and complexity in the interpretation allow some interesting trends to emerge. Figure 1 tries to model the cost of adding "standard-of-care" mutation analyses to the molecular testing of colorectal cancer (gradually completing a panel that would include *KRAS, NRAS, BRAF* and *PIK3Ca* mutation testing) and the level of complexity in the interpretation of the results. From this calculation one can see that a) NGS costs probably fall somewhere between laboratory-developed testing with technologies such as Sanger or Pyrosequencing (the most affordable) and the accumulation of FDA/CE IVD single-gene tests (the most expensive); but b) the most expensive is also the easiest to interpret, with NGS requiring some form of bioinformatics analysis, QC and curation of the results. This dichotomization of reagent cost versus interpretative complexity does not take into account other disadvantages of the NGS approach (for instance, a longer TAT with more hands-on technical time that can make the end product more expensive) and other technical advantages as well (such as the fact that, in terms of sample quantity, the level of multiplexing of current NGS technologies can reduce considerably the amount of DNA needed compared to individual gene testing). In this sense, it is sensible to predict that with full automation of sample processing and analysis, the costs of panel testing with NGS technologies can be affordable in routine healthcare.

How does NGS perform with routine clinical sample material?

This article is protected by copyright. All rights reserved

Accepted Article

We are beginning to experience that NGS technology, as it is available today, is less tolerant to suboptimal-quality DNA than the existing, more robust single-gene test protocols. This may be particularly challenging in laboratory environments such as those that are reference centres for many hospitals, where the control of pre-analytical variables is difficult or not possible. In addition, a significant degree of interest in mutation detection in peripheral blood (bringing the paradigm of "molecular monitoring" from leukaemia to solid tumors) will also require different technical approaches. In essence, implementation of NGS requires developing more robust protocols to become an everyday diagnostic tool applicable to all clinically available specimens.

To fully design new strategies for diagnostic validation of high-throughput technologies, it may be pertinent to see how other areas of medicine have faced similar challenges. The last few years have witnessed a revolution in oncology due to the advent of personalised medicine; this has also deeply transformed the way we practice pathology and molecular diagnostics. Oncologists have realised the unworkable long-term future of traditional clinical trial models, and have adopted adaptive trial designs [9] trying to answer multiple clinical questions at once. The molecular diagnostic equivalent to adaptive trials would require models where a "baseline" validation for key mutations is achieved, while criteria are laid out to detect and report other variants, or the same key mutations in a much lower frequency than usual are detected. These *adaptive validation designs* would require long-term monitoring of routine testing, and perhaps regular diagnostic confirmations by single gene test approaches, but would allow the utilization of NGS and other high-throughput technologies for diagnostics at an earlier stage.

***NGS – a broad-range detection technical approach*.**

This article is protected by copyright. All rights reserved

Accepted Article

While most commercially available NGS panels only look for single nucleotide variations and small insertion and deletions, the technology is showing great promise in the detection of other key genetic alterations of diagnostic interest, namely translocations in leukaemias, sarcomas, lymphomas, and other solid tumours such as NSCLC or gene copy number variation in breast cancer (*HER2* amplification), paediatric cancers (*N-MYC* amplification) and lymphomas (*C-MYC* amplification), among others. This promise of detection of genomic alterations at several levels, if fully achieved, will diminish the degree of technological complexity that we currently experience in our diagnostic laboratories, where multiple technologies such as RQ-PCR, Sanger sequencing, FISH and IHC can easily be applied to one single specimen today.

One of the most relevant aspects of the work by Tohill et al is the link between site of origin and genotype. Indeed, NGS studies will lead to the identification of novel variants of genomic alterations that have not been included in the existing trials or in retrospective clinical studies. This is important as we begin to understand that a similar genetic finding (for instance, the *BRAF*V600 mutation) holds different clinical and therapeutic value in different cancer types. For instance, while *BRAF*V600 mutations in melanoma open the avenue for treatment with vemurafenib monotherapy or in combination with MEK inhibitors, this is not the case in colorectal cancer, unless combined with anti-EGFR therapies. One cannot assume that the same therapies can be effective, as the mechanism of action of small molecule inhibitors is very sensitive to conformational changes in the protein that are cancer specific as well as to the surrounding feedback regulation loops and alternative signaling cascades that are driving tumor progression in different neoplasias.

This article is protected by copyright. All rights reserved

Accepted Article

While some agencies are already setting up criteria for these validations to take place [10], including key aspects in the NGS analytical process such as minimum coverage / depth of coverage, as well as key software parameters to demonstrate the accuracy and reproducibility of the bioinformatics process, there is still no unified model for the validation of NGS into routine diagnosis. Based on the parameters discussed in his paper, one could attempt to define a validation model based on case numbers, precision and LOD. However, the broad nature of NGS, the large number of possible targets to be detected and their level of clinical significance, and the increased sensitivity generated by massive parallel sequencing, makes a scientifically acceptable and cost effective modeling of these validations a very difficult endeavour. Hence, we feel that a more coordinated effort should occur to make the most of the promise of NGS in patient diagnosis and patient care. As such, the definition of validation guidelines by national colleges (RCPath, CAP or RCPA) or the establishment of international consortia for the validation of NGS in the diagnosis of solid tumors would lead to more robust and focused analytical laboratory processes, and to a much more efficient use of resources.

References

1. Tothill RW, Li J, Mileshkin L*, et al.* Massively-parallel sequencing assists the diagnosis and guided treatment of cancers of unknown primary. . *J Pathol.* 2013 DOI: 10.1002/path.4251.
2. Witkowski L, Lalonde E, Zhang J*, et al.* Familial rhabdoid tumour 'avant la lettre'--from

This article is protected by copyright. All rights reserved

Accepted Article

pathology review to exome sequencing and back again. *J Pathol.* 2013;**231**:35-43.

3. Catherwood MA, Maxwell P, Waugh DJ, *et al*. Validation of next generation sequencing technologies in comparison to current diagnostic gold standards for BRAF, EGFR and KRAS mutational analysis. McCourt CM, McArt DG, Mills K, *PLoS One* 2013;8(7):e69604.
4. Frampton GM, Fichtenholtz A, Otto GA, *et al*. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat Biotechnol.* 2013;31:1023-31.
5. Pritchard CC, Salipante SJ, Koehler K, *et al*. Validation and implementation of targeted capture and sequencing for the detection of actionable mutation, copy number variation, and gene rearrangement in clinical cancer specimens. *J Mol Diagn.* 2014;16:56-67.
6. Mattocks CJ, Morris MA, Matthijs G, et al. A standardized framework for the validation and verification of clinical molecular genetic tests. Eur J Hum Genet. 2010;18:1276-88.
7. Lindeman NI, Cagle PT, Beasley MB, *et al*. Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. J Mol Diagn. 2013;15:415-53.
8. André F, Bachelot T, Commo F, *et al*. Comparative genomic hybridisation array and DNA sequencing to direct treatment of metastatic breast cancer: a multicentre, prospective trial (SAFIR01/UNICANCER). Lancet Oncol. 2014. pii: S1470-2045(13)70611-9. DOI: 10.1016/S1470-2045(13)70611-9.
9. Berry DA. Adaptive clinical trials in oncology. Nat Rev Clin Oncol. 2011 Nov 8;9(4):199-207.
10. http://www.wadsworth.org/labcert/TestApproval/forms/NextGenSeq_ONCO_Guidelines.pdf

This article is protected by copyright. All rights reserved

Accepted Article

TABLES

Table 1 – NGS list of considerations prior to diagnostic validation

Before NGS is validated for molecular diagnostic purposes
Should every target be validated?
Should we calculate sensitivity / specificity / precision / accuracy for every target? And for every sample type?
How do we validate the bioinformatics curation of the results?
It the turnaround time (TAT) competitive?
What is the minimum number of targets to make NGS cost-efficient in diagnostics?
How does NGS perform with routine clinical sample material?

This article is protected by copyright. All rights reserved

Accepted Article

Table 2 – The process of NGS testing

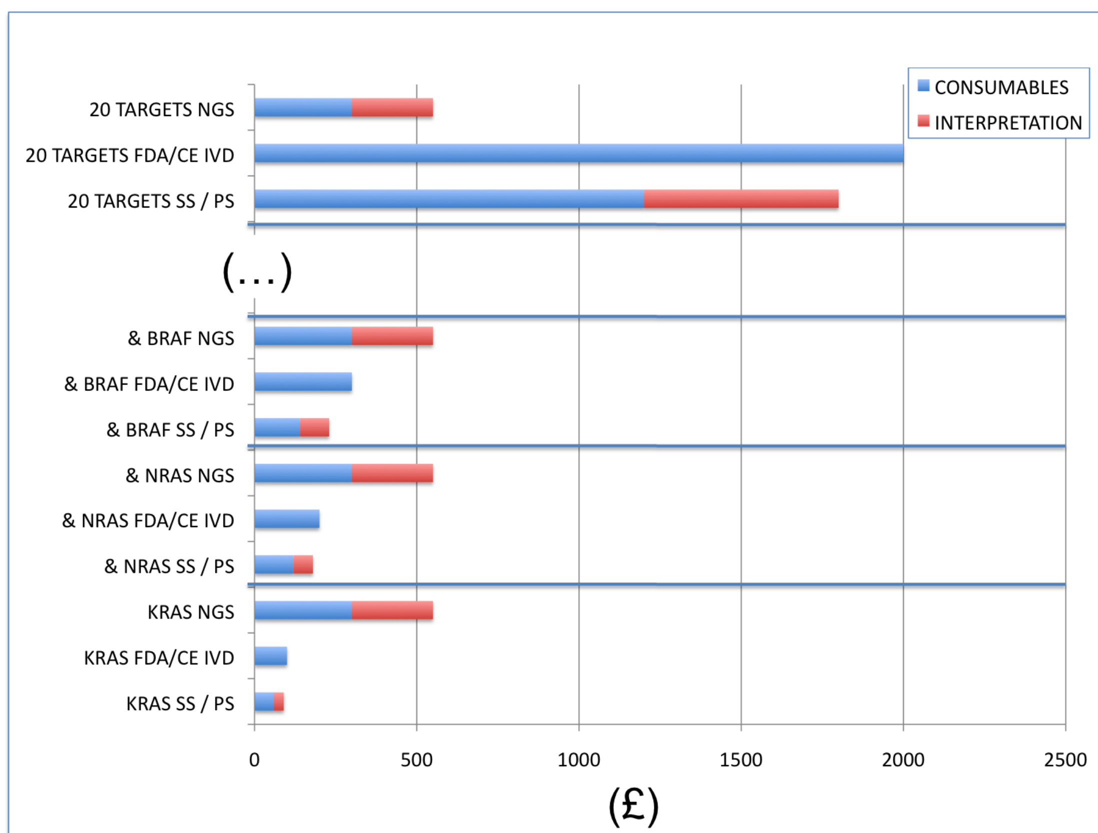
BASIC NGS ANALYTICAL PROCESS	
RECEIPT AND ASSESMENT OF FFPE TISSUE	1-2 DAYS
MICRODISSECTION	
DNA EXTRACTION (OVERNIGHT)	
LIBRARY PREPARATION	2-5 DAYS
(ENRICHMENT IF CAPTURE APPROACHES)	
DNA SEQUENCING	
BIOINFORMATICS ANALYSIS	1-3 DAYS
PREPARATION /AUTHORISATION OF REPORT	

This article is protected by copyright. All rights reserved

Accepted Article

Legend to Figure 1 –

Cost projection (consumables and pathologist's time for interpretation) of standard of care gene mutation analysis with 3 types of technology (SS/PS = Sanger Sequeincing / Pyrosequencing, typically homebrew; IVD – in-vitro diagnostics; NGS – Next Generation Sequencing). The model takes the following assumptions: SS/PS homebrew has a baseline cost of £20 per reaction, and an increasing complexity of interpretation as new reactions are added; IVD has a baseline cost of £100 per test, but the interpretation is in-built in the system; NGS has a similar cost of consumables and interpretation (for standard-of-care gene mutations), regardless of the number of standard-of-care genes



This article is protected by copyright. All rights reserved