

Variations in pre-analytical FFPE sample processing and bioinformatics: challenges for next generation molecular diagnostic testing in clinical pathology

Advances in cellular pathology techniques will improve diagnostic medicine. However, such improvements have to overcome many challenges including variations in pre-analytical sample processing, bioinformatics data analysis and clinical interpretation of data. In order to resolve such challenges, bioinformatics needs to become more tightly coupled to the experimental methodology development.

by Dr Rifat Hamoudi, Dr Joshua Kapp, Sevgi Umur and Michael Gandy

Introduction

Molecular diagnostics within cellular pathology have been performed since the late 1990s and have developed to include a range of techniques including short tandem repeat (STR) identity analysis, classification of tumours and clonality determinations in hematopathology. More recently, with the introduction of qPCR and more recently of next generation sequencing (NGS) as shown in Figure 1, precision medicine testing for targeted therapies has rapidly gained access to daily practice and become a challenge for molecular biologists and pathologists to provide the most accurate and

relevant information. As part of this testing process we discuss two major challenges which have developed, these are:

- Firstly, pre-analytical processing of formalin-fixed paraffin-embedded (FFPE) tissue, has shown to be a critical determinant in the accuracy of downstream molecular testing in specialities such as mutational screening for targeted therapies.
- Secondly, bioinformatics has become a bottleneck in data processing and interpretation, with the processing, analysis and reporting of the data shown variability between different laboratories.

This article looks to raise the awareness of these issues and presents possible areas for consideration to aid in their resolution.

Variation in pre-analytical sample processing of FFPE samples may lead to discrepancies in mutational testing of actionable genes

Within cellular pathology, the majority of molecular diagnostic clinical sample testing is now carried out on FFPE samples. Generally the tissue is screened using hematoxylin and eosin stained sections to estimate the tumour content before the preparation process of material for subsequent molecular testing, as shown in Figure 2.

Recent studies have shown that variations in pre-analytical processing of samples lead to discrepancies in downstream molecular diagnostic testing [1–3]. The variations using singleplex mutational screening were largely due to the DNA extraction system used [2, 3], quantitation using spectrophotometry and training of laboratory staff as one study showed that pre-analytical variation was significant even among experienced laboratories [3]. In addition both DNA quantitation and integrity measurements play important roles in the accuracy of downstream multiplex testing using NGS.

In order to resolve some of those issues it is important to include control series of diagnostic samples, prepared according to the diagnostic operating procedures of the laboratory with a variety of known mutations comprising missense mutations, simple and complex deletions and insertions. Assay control using known representative DNA samples from the FFPE tissue is also essential to ensure that the process of DNA extraction, quantitation and integrity measurements are performed correctly and consistently. This is important as DNA quality has a major effect on NGS performance, i.e. poor quality DNA causes a higher error rate [3].

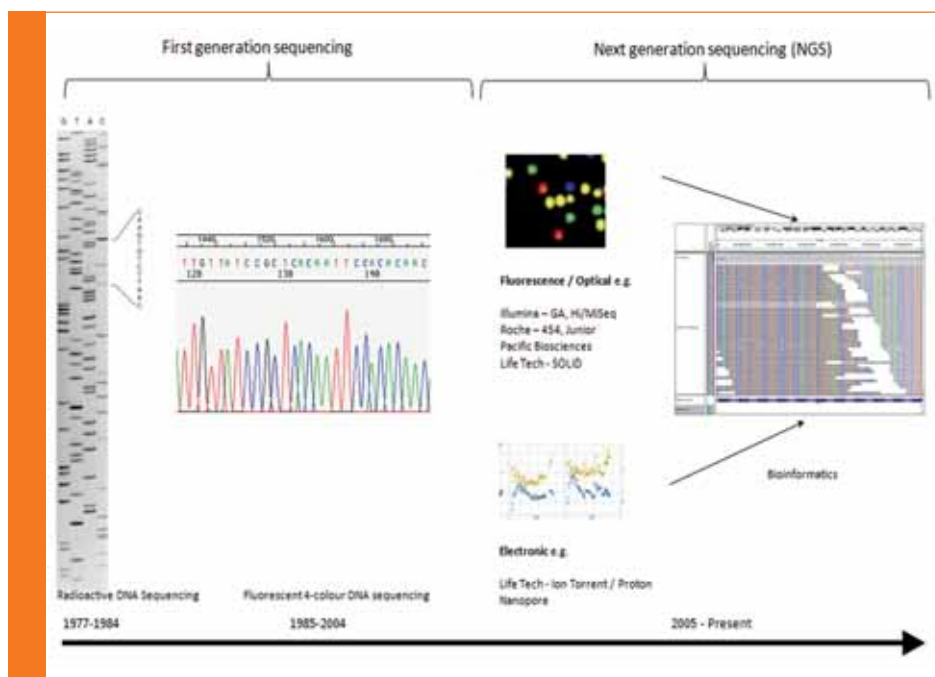


Figure 1. Historical development of sequencing technology.

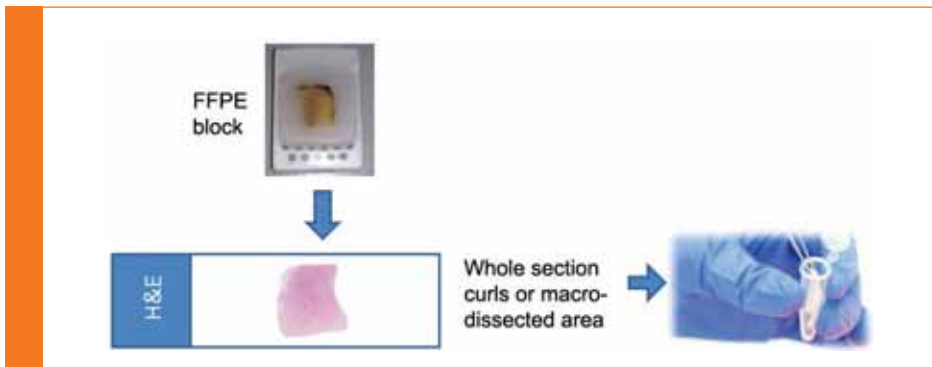


Figure 2. Formalin-fixed paraffin-embedded (FFPE) tissue preparation for molecular pathology testing.

In addition, differences in quantitation measurements need to be accounted for, since the different instruments used have different ways of measuring the concentration of DNA. For example, variations can be seen between systems such as Nanodrop spectrophotometry and Qubit fluorometry. Measurement of DNA integrity is also important and most labs use assays such as BIOMED [4, 5] or qPCR as the 'gold standard' measure.

Also European external quality assurance (EQA) programmes for mutation detection of solid tumours such as European Society for Pathology (ESP, www.esp-pathology.org), European Molecular Genetics Quality Network (EMQN, www.emqn.org), and United Kingdom National External Quality Assessment Scheme UK NEQAS for Molecular Pathology (www.ukneqas.org.uk and www.ukneqas-molgen.org.uk) may consider including pre-analytical (e.g. pre-PCR) component in their assessment for mutation detection from FFPE samples.

Discrepancies in variant-calling pipelines and high-throughput sequencing clinical interpretation

Most diseases such as cancer and inherited diseases are driven by genomic alterations. Recent advances in high-throughput sequencing technologies have enabled the identification of somatic mutations at very high resolution. However, accurate somatic mutation-calling using high-throughput sequence data remains one of the major challenges in genomics. For somatic mutation-calling, one looks for a site in which a variant allele exists in the tumour sample but not in the normal sample. Even with the sequence data from a normal sample, variant-calling in high-throughput sequencing data is challenging due to the multiple potential sources of errors. For example, artefacts occurring during PCR amplification or targeted capture (e.g. exome-capture), machine sequencing errors, and incorrect local alignments of reads are all well documented sources of error [6–8]. Tumour heterogeneity and

normal contamination contribute additional challenges for the tumour samples [9].

Various studies have shown low concordance between different variant callers and bioinformatics analysis pipelines. Wang *et al.* [10] compared six variant callers on whole exome sequencing melanoma sample and matched blood of 18 lung tumour–normal pairs and seven lung cancer cell lines carried out on the Illumina HiSeq 2000. The results showed discordance between the six variant callers, and the top two performing callers could only detect 86% and 71% of validated mutations respectively. O'Rawe *et al.* [11] compared the analysis of five different Illumina alignment and variant-calling pipelines on 15 exome sequencing data carried out using Illumina HiSeq 2000 and Agilent SureSelect version 2 capture kit at 120X mean coverage. Results showed variant-calling concordance of 57.4% between the five different Illumina pipelines across all 15 exomes with the authors urging more caution when analysing individual genomes in genomic medicine. In addition, comparison of the two most prominent cancer genome sequencing databases; catalogue of somatic mutations in cancer (COSMIC) [12] and Cancer Cell Line Encyclopaedia (CCLE) [13] revealed marked discrepancies in the detection of missense mutations in identical cell lines (57.4% conformity), where the main reason for such discrepancy is inadequate sequencing of GC-rich areas of the exome [14].

In addition to the above, various studies have shown discrepancies in the interpretation of genomic data between the clinician and diagnostic laboratory. Shashi *et al.* [15] tried to follow up the results of 93 patients who underwent exome sequencing. They investigated how the clinical interpretation of the lab results changed the diagnosis and its conformity with it. Overall, the results showed that in 25% of patients (24/93), exome sequencing showed a positive result

and in 80% (19/24) of cases, the clinicians agreed with the molecular diagnosis of the lab. However, in 20% of patients reported to be positive by the diagnostic lab, the clinicians thought that the suggested molecular diagnosis was not correct. In addition, 5% of patients that were considered negative by the exome lab or had a lower confidence diagnosis, were eventually found to be positive when the exome data was reviewed by clinicians. In summary the results showed 20% false positives and 5% false negatives when comparing the interpretation of genomic data between different healthcare staff.

However, it is worth noting that all the above studies used samples with high molecular weight DNA from cell lines, fresh frozen tissue or blood and carrying out the same studies above using FFPE samples has the potential to lead to further discrepancies due to the degraded DNA inherent to those samples increases the variation at the pre-analytical steps resulting in downstream discrepancies in mutational profiling. This creates a big challenge in the development of bioinformatics pipelines required to produce consistent clinically reliable data.

One way to resolve some of the bioinformatics related issues is to exchange the raw datasets between laboratories that preferentially use different software as part of the software validation process to establish the ability of the various laboratories to detect identical gene mutations. In addition, new software updates need to be validated by analysis of prior NGS datasets covering simple and complex mutations. Finally, raw NGS datasets need to be included in EQA programmes as *in silico* assessment.

Conclusion

Although the above discussion very briefly surveys the current landscape in cellular pathology, the future of molecular diagnostics will undoubtedly develop to include integrated RNA expression analysis, DNA amplification and epigenetics. Each methodology will have its own idiosyncrasies and will require the development of new clinically validated bioinformatics pipeline. Additionally, the need for a novel bioinformatics system to support integrative analysis will become essential. Although previously attempted [16], new systems need to be developed to support integrative high-throughput sequencing analysis.

However, before novel bioinformatics software solutions can be devised for big data, concerns about bioinformatics software development need to be addressed. A potential starting point to

address this is via supporting new bioinformatics courses that use software engineering, computer programming and mathematical modelling of biological complexity at their core, supporting the education of future bioinformaticians in the art of bioinformatics software development. This will help support a change in the current paradigm where much of the current bespoke bioinformatics software today has been developed by local institutions in relative isolation, often in conjunction within the framework of a specialist area experimental research program [17].

The future landscape highly likely see the validation of wet chemistries (laboratory and clinical based) and dry (computational based) experiments carried out in more tightly coupled format than is currently performed, supporting clinical product development in the commercial market. Also, the future will see more focus on the development of more efficient adaptive algorithms that address the clinical questions, leading to faster analysis and improving the clarity in the interpretation of the data.

In conclusion, within cellular pathology the incremental development of pre-analytical processing from FFPE samples coupled with more efficient adaptive bioinformatics algorithms implementation are key areas of focus and crucial to the further advancement of next generation molecular pathology.

References

1. Carrick DM, Mehaffey MG, Sachs MC, Altekruze S, *et al.* Robustness of Next Generation Sequencing on older formalin-fixed paraffin-embedded tissue. *PLoS One* 2015; 10: e0127353.
2. Heydt C, Fassunke J, Kunstlinger H, Ihle MA, *et al.*

- Comparison of pre-analytical FFPE sample preparation methods and their impact on massively parallel sequencing in routine diagnostics. *PLoS One* 2014; 9: e104566.
3. Kapp JR, Diss T, Spicer J, Gandy M, *et al.* Variation in pre-PCR processing of FFPE samples leads to discrepancies in BRAF and EGFR mutation detection: a diagnostic RING trial. *J Clin Pathol.* 2015; 68: 111–118.
4. Johnson NA, Hamoudi RA, Ichimura K, Liu L, *et al.* Application of array CGH on archival formalin-fixed paraffin-embedded tissues including small numbers of microdissected cells. *Lab Invest.* 2006; 86: 968–978.
5. van Dongen JJ, Langerak AW, Bruggemann M, Evans PA, *et al.* Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98–3936. *Leukemia* 2003; 17: 2257–2317.
6. Meacham F, Boffelli D, Dhabhi J, Martin DI, *et al.* Identification and correction of systematic error in high-throughput sequence data. *BMC Bioinformatics* 2011; 12: 451.
7. Nakamura K, Oshima T, Morimoto T, Ikeda S, *et al.* Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res.* 2011; 39: e90.
8. Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. *Nat Rev Genet.* 2011; 12: 443–451.
9. Gerlinger M, Rowan AJ, Horswell S, Larkin J, *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med.* 2012; 366: 883–892.
10. Wang Q, Jia P, Li F, Chen H, *et al.* Detecting somatic point mutations in cancer genome sequencing data: a comparison of mutation callers. *Genome Med.* 2013; 5: 91.
11. O’Rawe J, Jiang T, Sun G, Wu Y, *et al.* Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. *Genome Med.* 2013; 5: 28.

12. Forbes SA, Beare D, Gunasekaran P, Leung K, *et al.* COSMIC: exploring the world’s knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* 2015; 43: D805–D811.
13. Barretina J, Caponigro G, Stransky N, Venkatesan K, *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012; 483: 603–607.
14. Hudson AM, Yates T, Li Y, Trotter EW, *et al.* Discrepancies in cancer genomic sequencing highlight opportunities for driver mutation discovery. *Cancer Res.* 2014; 74: 6390–6396.
15. Shashi V, McConkie-Rosell A, Schoch K, Kasturi V, *et al.* Practical considerations in the clinical application of whole-exome sequencing. *Clin Genet.* 2015; doi: 10.1111/cge.12569.
16. Watkins AJ, Hamoudi RA, Zeng N, Yan Q, *et al.* An integrated genomic and expression analysis of 7q deletion in splenic marginal zone lymphoma. *PLoS One* 2012; 7: e44997.
17. Prins P, de Ligt J, Tarasov A, Jansen RC, *et al.* Toward effective software solutions for big biology. *Nat Biotechnol.* 2015; 33: 686–687.

The authors

Rifat Hamoudi*¹ PhD, Joshua Kapp¹ MBBS, Sevgi Umur² BSc and Michael Gandy³ MSc
¹Division of Surgery and Interventional Science, University College London, London, UK
²Genomymous Sciences, Küçükçekirmez, Defne Sokak, Flora Residence Istanbul, Turkey
³Health Services Laboratories, 60 Whitfield Street, London, UK

*Corresponding author
 E-mail: r.hamoudi@ucl.ac.uk

EVENT PREVIEW

POCT and preanalytics to be the themes of Labquality Days 2016



The Labquality Days Congress will be held at the Messukeskus Expo and Convention Centre in Helsinki on 11th-12th of February 2016. Labquality Days is one of the largest annual congresses in Scandinavia focused on quality and laboratory medicine. The congress inspires clinical chemistry, laboratory medicine professionals, researchers, healthcare experts, users of point-of-care devices, medical staff working with quality issues, managers and higher level personnel administration of social- and or healthcare sectors. The 2016 congress themes are now announced: Point-of-Care Testing (POCT) and preanalytics. POCT has already a major role in healthcare workflow. Test sensitivity or specificity, price, speed and patient convenience are some heavily discussed topics in scientific meetings. In preanalytics, various disciplines such as microbiology, clinical chemistry and hematology have their own characteristic variables. Individual analyses have some unique factors that should also be taken into account in order to obtain reliable results. Labquality Days will bring together leading international speakers and opinion leaders. The programme consists of scientific lectures and panel discussions. During the congress participants have the opportunity to meet colleagues, share ideas and experience the vast clinical laboratory exhibition.