Chapter 5
DNA-Based Sequencing Assays

**Pasquale Pisapia, Miriam Cieri, Francesco Pepe,
Umberto Malapelle, and Giancarlo Troncone**

Abbreviations

BRAF	v-raf murine sarcoma viral oncogene homolog B
DNA	Deoxyribonucleic acid
EGFR	Epidermal growth factor receptor
IVD	In vitro diagnostic
KIT	KIT proto-oncogene receptor tyrosine kinase
KRAS	Kirsten rat sarcoma viral oncogene homolog
LDT	Laboratory developed tests
LOD	Limit of detection
NGS	Next-generation sequencing
NRAS	Neuroblastoma RAS viral oncogene
PDGFRA	Platelet-derived growth factor receptor A
PPA	Positive percentage agreement
PPV	Positive predictive value
RT-PCR	Real-time polymerase chain reaction
TAT	Turnaround time

P. Pisapia · M. Cieri · F. Pepe · U. Malapelle · G. Troncone (✉)
Department of Public Health, University of Naples "Federico II",
Naples, Italy
e-mail: francesco.pepe4@unina.it; umberto.malapelle@unina.it;
giancarlo.troncone@unina.it

Key Terminology

|	
DNA library	Collection of DNA fragments that are captured, barcoded, and clonally amplified, prior to sequencing on NGS platforms.
Gene panel	Representative gene regions covered by a sequencing assay.
Laboratory developed tests	Test designed, developed, and adapted in-house after validation.
Limit of detection	Corresponds to the analytical sensitivity of a given NGS assay, reflecting the lowest amount of analyte which can be reliably detected.
Molecular cytopathology	Discipline of cytopathology based on the integration of morphologic changes with the genomic alterations/molecular features underlying the development, progression, and prognosis of neoplastic diseases.
Next-generation sequencing	High-throughput molecular platform that allows sequencing multiple gene sequences in parallel and interrogating various genetic alterations for multiple patients in a single run.
Personalized medicine	Cancer therapy based on the specific molecular alterations of a patient's tumor.
Pyrosequencing	"sequencing by synthesis"-based technology, in which the sequential incorporation of nucleotides is identified by the

	detection of a released pyrophosphate.
Reads	DNA fragments that are sequenced by a NGS platform during a run.
Real-Time PCR	PCR-based assay that detects and quantifies in "real time" the amplification of a given DNA target using specific fluorescent probes.
Reference range	The interval between the upper and lower concentrations of analyte in the sample for which a suitable level of precision, accuracy, and linearity has been demonstrated
Sanger sequencing	Standard sequencing technology based on the incorporation of chain-terminating dideoxynucleotides (usually fluorochrome labeled) during the process of sequencing by DNA polymerase.
Turnaround time	Time required to analyze a sample and deliver a test result from when the sample is accessioned in the laboratory.
Validation	Procedure that defines the performance parameters of an assay, such as sensitivity, specificity, accuracy, precision, detection limit, range, and limits of quantitation of a novel methodology prior to clinical implementation.

Key Points

- Molecular cytopathology plays a key role in clinical diagnostics, prognostication, and the selection of patients for targeted treatment
- Modern cytopathologists need to be familiar with molecular techniques to appropriately triage specimens for molecular testing
- In comparison to histologic material, cytology specimens often provide better-quality DNA
- Most DNA-based assays including NGS can be successfully applied to cytology specimens
- In order to improve NGS laboratory workflow, it is important to create a gene panel to cover relevant hotspot targets with a defined cost
- In-house validation of each new diagnostic methodology implemented in routine practice is required, even when commercially available and validated for in vitro diagnostic use

In the last 10 years, the landscape of *Personalized Medicine* has included the contribution of *Molecular Cytopathology*, in particular for advanced stage patients with solid tumors. Since these patients are not candidates for surgical resection, a concurrent histology specimen is not always available [1, 2]. Therefore, in order to be a knowledgeable partner in diagnostic and predictive approaches to cancer therapy, the modern cytopathologist needs to be familiar with the basic principles and some of the more advanced molecular techniques used in clinical practice [3, 4].

Cytology samples (Fig. 5.1) provide high-quality DNA, sufficient for a wide array of DNA-based sequencing assays, including next-generation sequencing (NGS) [5]. This novel high-throughput technology represents an evolution of conventional DNA sequencing methodologies, such as Sanger sequencing and pyrosequencing.
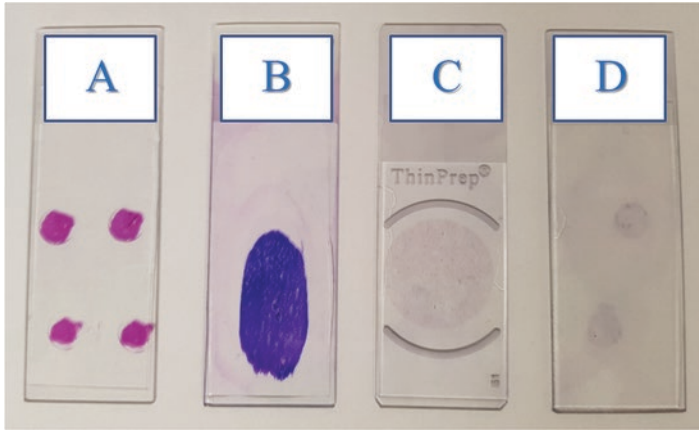
FIGURE 5.1 Examples of different cytologic preparations commonly used for molecular assays: (**A**) cell block; (**B**) direct smear; (**C**) liquid-based cytology; (**D**) cytospin

Sanger sequencing has long been the gold standard for the identification of point mutations, deletions, and small insertions [6, 7]. In this method, a chemically modified nucleotide (dideoxynucleotide) terminates the extension of the DNA strand at the point of incorporation. This results in a mixture of DNA fragments of varying lengths. Each dideoxynucleotide, (A, T, C, or G) is labeled with a different fluorescent dye (dye terminator). The newly synthesized and labeled DNA fragments are sequentially separated by size through capillary gel electrophoresis. The fluorescence is detected by an automated sequence analyzer, and the order of nucleotides (base calling) in the target DNA is visualized as a sequence electropherogram [7, 8]. Although Sanger sequencing was the method first employed in most clinical pathology laboratories, its low sensitivity (around 20% of mutant alleles) limits its application in low tumor content samples, in which the tumor often constitutes a minority of the mixed cell population present. Thus, Sanger sequencing frequently requires tumor enrichment by microdissection prior to analysis to avoid false-negative results [9]. Although low throughput,

TABLE 5.1 Sanger sequencing: principal advantages and disadvantages

Advantages	Disadvantages
Gold standard for single nucleotide polymorphisms	Single-gene testing, low-throughput technique
Robust technology featuring high specificity	Low sensitivity (LOD 5–20% mutant allele)
Enables the detection of both common and uncommon gene mutations	High turnaround time
	Several steps post PCR
Simple data interpretation	Uncorporated fluorescent nucleotides may appear very
Suitable on FFPE samples	prominent (dye blobs)
	Generates false-negative results on low tumor fraction samples

Sanger sequencing is a robust technology, suitable for analyzing complex genomic regions featuring combined deletion and insertions (Table 5.1).

Pyrosequencing is another method of DNA sequencing by synthesis and is a valid alternative to Sanger sequencing. It relies on the detection of a pyrophosphate released during the DNA polymerase reaction with an enzymatic cascade resulting in the production of visible light [10]. This is converted in analog signal as a peak in a pyrogram. Pyrosequencing provides higher sensitivity (around 5% of mutated allele) than Sanger sequencing, but its error rate (1.07%) is not negligible [11]. When a heterozygous mutation is identified by direct sequencing or by pyrosequencing, both mutant and wild-type alleles are seen on the sequencing electropherograms and on the pyrograms, respectively [10, 11].

With the increase in the number of predictive and prognostic biomarker testing needed for patient management, there is a growing need for high-throughput sequencing technology with the capability of evaluating multiple genes simultaneously. A suitable and flexible multigene testing approach to evaluating known somatic point mutations is by the Sequenom MassARRAY®. This genotyping platform is based on the matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry and can provide

customized genotyping assays to analyze allele-specific primer extension products [12]. The basic principle underlying this assay is that mutant and wild-type alleles for a given point mutation produce single-allele base extension reaction products of a mass that is specific to the sequence of the product. Mutation calls are based on the mass differences between the wild-type product and the mutant products as resolved by MALDI-TOF mass spectrometry [12].

Compared to conventional sequencing technologies, next-generation sequencing (NGS) offers high analytic sensitivity together with a high clinical sensitivity. Analytic sensitivity (also known as allelic fraction) is defined as the ability of a mutational assay to identify an alteration in a background of wild-type alleles. Clinical sensitivity covers the spectrum of possible alterations that can be identified by any given assay [13]. NGS exploits a massively parallel sequencing technology, which increases sequencing throughput from hundreds of thousands to millions of sequences (reads) and enables simultaneous analyses of different gene targets for multiple patients in each run [14, 15]. The balance between analytic and clinical sensitivity seen in NGS, together with the minimal amounts of input DNA required, makes this technology ideal for application in cytology samples. The increasing use of NGS in combination with advanced tumor sampling techniques using novel bronchoscopic/endoscopic approaches makes the practice of cytopathology an attractive field in the realm of molecular medicine [16, 17]. A key advantage of NGS over more targeted sequencing technologies is the opportunity to evaluate biomarkers in novel genes of potential clinical interest, in addition to standard of care testing, and thereby facilitate enrollment of patients in clinical trials [18–20].

The principal advantages and limitations of NGS are listed in Table 5.2 [5, 15, 18–24].

A variety of NGS platforms are available for clinical use. Despite the availability of different platforms, the NGS workflow is characterized by four principal steps: (1) DNA library generation, (2) single fragment clonal amplification, (3) massive parallel sequencing, and (4) data analysis [16, 21, 25] (Fig. 5.2).

TABLE 5.2 Next-generation sequencing: principal advantages and disadvantages

Advantages	Disadvantages
Suitable on paucicellular samples	Extensive in-house validation
High throughput	
Open and flexible technology	Need of orthogonal techniques to validate the assay and refine borderline results
Analysis for multiple patients simultaneously in a single run	
Evaluation of mutational assessment for prediction, prognostic, and clinical trials purposes	Challenges in bioinformatics interpretation
	Validation for different sample preparation
Wide reference range	Short average read lengths
Shorter turnaround time and improved laboratory cost-effectiveness	Costs and reimbursement
Choice of panel	
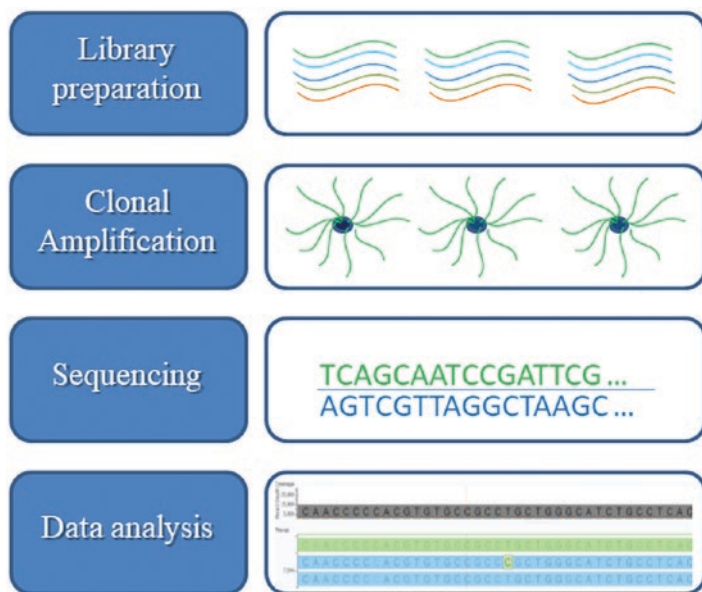
FIGURE 5.2 Schematic representation of the four steps of the NGS workflow, including DNA library preparation, single fragment clonal amplification, massive parallel sequencing, and data analysis

The DNA input required to generate the library is dependent on the target gene selection. The Illumina[TM] platforms (San Diego, CA, USA) utilize a hybridization-based capture system and require a DNA input ranging from 50 to 250 ng and 24–72 h for processing the sequencing data [22]. Recent advances in library preparation have enabled a reduction in the required input DNA, and Illumina validated protocols can be optimized to analyze 10–100 ng of DNA [26]. Alternatively, the IonTorrent platforms[TM] (Life Technologies, Carlsbad, CA, USA) utilize an amplicon-based technology. Multiple primer pairs are employed to select target gene regions by PCR, which requires as little as 10 ng (or even less) DNA input and only 1–3 h to generate the sequence results [23]. Another NGS platform, the GeneReader NGS System[TM] (Qiagen, Hilden, Germany), more recently became available. This platform requires at least 40 ng of DNA, adopts a hybridization-based library preparation methodology, and requires a relatively long analysis time (approximately 30 h) [27].

Clonal amplification is the second step in the NGS workflow. To enhance the chemical signal in the subsequent sequencing reaction, each single fragment of the library needs to be clonally expanded in hundreds of thousands of copies [15]. On the Illumina platform, clonal amplification takes place on a solid support of a flat glass microfluidic channel (flow cell) by the so-called bridge amplification [22], whereas the Ion Torrent and GeneReader platforms carry out clonal amplification by emulsion PCR on beads [5, 21, 27].

The third step in the NGS workflow is the massive parallel sequencing with generation of hundreds of thousands to millions of reads in parallel for each run [14, 27]. The differences among the most commonly adopted platforms are highlighted in Table 5.3. Despite the differences in DNA input requirements, the run times, read lengths, and costs per sample, the two most popular bench-top sequencing platforms (Illumina and Ion Torrent) produce comparable results [16].

Finally, sequencing data are analyzed by using a combination of software pipelines (Fig. 5.3) [27, 28]. This process requires four major steps: base calling, read alignment, variant identification, and variant annotation [29, 30].

TABLE 5.3 Difference between the commonly used NGS platforms in clinical laboratories

	Ion Torrent PGM	Illumina	GeneReader
Input [DNA] ng to generate library	10 or less	10–100	40
DNA library generation	Amplicon-based system	Hybridization-based capture system	Hybridization-based capture system
Clonal amplification	Emulsion PCR	Bridge PCR	Emulsion PCR
Sequencing signal	pH change	Reversible dye terminators	Reversible dye terminators
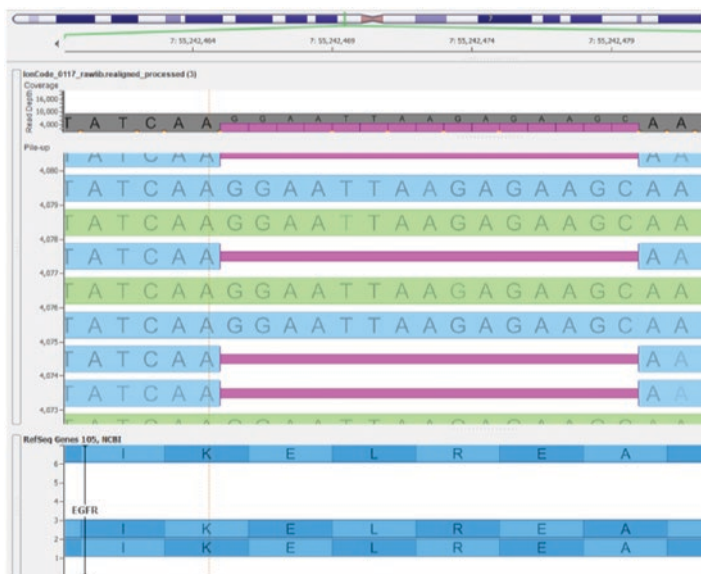
FIGURE 5.3 *EGFR* mutation analysis by NGS. Read alignment visualization of Golden Helix GenomeBrowse v.2.0.7 (Bozeman, MT, USA) software showing an epidermal growth factor receptor (EGFR) exon 19 deletion (p.E746_A750delELREA)

The combination of informatics tools used for processing, aligning, and detecting variants in NGS data is commonly referred to as the bioinformatics pipeline. This process requires careful optimization at the time of validation to ensure that a variant call is effectively present in the sequence as well as continued quality control, as bioinformatics is constantly evolving. The necessity of validation of NGS technologies prior to clinical implementation cannot be overemphasized. Validation includes the identification of positive percentage agreement (PPA) and positive predictive value (PPV), the reproducibility of variant detection, the determination of the reference range, limits of detection (LOD), clinical and analytical sensitivity and specificity, and if appropriate, the validation of bioinformatics pipelines, and other parameters [31].

NGS is a powerful and versatile technique. A variety of gene panels is commercially available and can be classified in four distinct groups, as summarized in Table 5.4 [31, 32].

The versatility of NGS lies in its ability to use custom panels to improve analytical performance and laboratory cost-effectiveness [3, 33]. Although it is widely held that NGS is an expensive technique, our experience with the commercially

TABLE 5.4 Examples of gene panels

Panels covering up to 10–15 actionable genes	Clinically relevant genes (narrow panels), these panels represent a viable alternative to RT-PCR assays
Panels covering up to 50 genes	Target both actionable and potentially clinically relevant genes, the assessment of which may affect eligibility for clinical trials
Panels covering up to 150 genes	Extensively targeting the common and uncommon driver genes of specific cancer type
Panels covering up to 400 genes	Essentially cover most of the known cancer-related genes. Such panels are also used to assess the so-called tumor mutational burden which may affect eligibility for cancer immunotherapy

available AmpliSeq Colon and Lung Cancer Panel, which covers 22 genes involved in colon and lung cancer, showed that the consumable cost is only €196 ($238) per sample [25]. Moreover, the cost per sample could be even reduced to €98 ($119) by the use of a narrow gene panel targeting 568 clinically relevant mutations in 6 genes (*EGFR*, *KRAS*, *NRAS*, *BRAF*, *KIT*, and *PDGFRA*) [33].

In summary, NGS-based assays on routine cytology samples have the potential to improve patient care through diagnostic, prognostic, and predictive biomarker assessment. The basic principles of NGS described in this chapter underscore the need of a new generation of molecular cytopathologists [34–36].

References

1. Clark DP. Seize the opportunity: underutilization of fine-needle aspiration biopsy to inform targeted cancer therapy decisions. Cancer. 2009;117:289–97.
2. Bellevicine C, Malapelle U, Vigliar E, et al. How to prepare cytological samples for molecular testing. J Clin Pathol. 2017;70:819–26.
3. Schmitt FC, Longatto-Filho A, Valent A, et al. Molecular techniques in cytopathology practice. J Clin Pathol. 2008;61:258–67.
4. Bellevicine C, Vita GD, Malapelle U, et al. Applications and limitations of oncogene mutation testing in clinical cytopathology. Semin Diagn Pathol. 2013;30:284–97.
5. Malapelle U, Mayo-de-Las-Casas C, Molina-Vila MA, et al. Consistency and reproducibility of next-generation sequencing and other multigene mutational assays: a worldwide ring trial study on quantitative cytological molecular reference specimens. Cancer. 2017;125:615–26.
6. McCourt CM, McArt DG, Mills K, et al. Validation of next generation sequencing technologies in comparison to current diagnostic gold standards for BRAF, EGFR and KRAS mutational analysis. PLoS One. 2013;8:e69604.
7. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA. 1977;74:5463–7.

8. Smith LM, Sanders JZ, Kaiser RJ, et al. Fluorescence detection in automated DNA sequence analysis. Nature. 1986;321:674–9.

9. Aisner DL, Rumery MD, Merrick DT, et al. Do more with less: tips and techniques for maximizing small biopsy and cytology specimens for molecular and ancillary testing: the University of Colorado experience. Arch Pathol Lab Med. 2016;140:1206–20.

10. Ronaghi M, Uhlén M, Nyrén P. A sequencing method based on real-time pyrophosphate. Science. 1998;363:281, 365.

11. Gilles A, Meglécz E, Pech N, et al. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. BMC Genomics. 2011;12:245.

12. Alì G, Proietti A, Pelliccioni S, et al. ALK rearrangement in a large series of consecutive non-small cell lung cancers: comparison between a new immunohistochemical approach and fluorescence in situ hybridization for the screening of patients eligible for crizotinib treatment. Arch Pathol Lab Med. 2014;138:1449–58.

13. Aisner DL, Sams SB. The role of cytology specimens in molecular testing of solid tumors: techniques, limitations, and opportunities. Diagn Cytopathol. 2012;40:511–24.

14. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008;456:53–9.

15. Rothberg JM, Hinz W, Rearick TM, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011;475:348–52.

16. Vigliar E, Malapelle U, de Luca C, et al. Challenges and opportunities of next-generation sequencing: a cytopathologist's perspective. Cytopathology. 2015;26:271–83.

17. Roy-Chowdhuri S, Stewart J. Preanalytic variables in cytology: lessons learned from next-generation sequencing-the MD Anderson experience. Arch Pathol Lab Med. 2016;140:1191–9. [Epub ahead of print].

18. Rizzo JM, Buck MJ. Key principles and clinical applications of "next-generation" DNA sequencing. Cancer Prev Res (Phila). 2012;5:887–900.

19. Malapelle U, Vigliar E, Sgariglia R, et al. Ion Torrent next-generation sequencing for routine identification of clinically relevant mutations in colorectal cancer patients. J Clin Pathol. 2015;68:64–8.

20. Malapelle U, Pisapia P, Sgariglia R, et al. Less frequently mutated genes in colorectal cancer: evidences from next-generation sequencing of 653 routine cases. J Clin Pathol. 2016;69:767–71.
21. Li J, Batcha AM, Grüning B, et al. An NGS workflow blueprint for DNA sequencing data and its application in individualized molecular oncology. Cancer Inform. 2016;14:87–107.
22. Mardis ER. Next-generation sequencing platforms. Ann Rev Anal Chem (Palo Alto Calif). 2013;6:287–303.
23. Beadling C, Neff TL, Heinrich MC, et al. Combining highly multiplexed PCR with semiconductor-based sequencing for rapid cancer genotyping. J Mol Diagn. 2013;15:171–6.
24. Rothberg JM, Leamon JH. The development and impact of 454 sequencing. Nat Biotechnol. 2008;26:1117–24.
25. Clinical and Laboratory Standards Institute. Nucleic acid sequencing methods in diagnostic laboratory medicine; approved guideline, CLSI document MM09-A2. 2nd ed. Wayne: Clinical and Laboratory Standards Institute; 2014.
26. Middleton G, Crack LR, Popat S, et al. The National Lung Matrix Trial: translating the biology of stratification in advanced non-small-cell lung cancer. Ann Oncol. 2015;26:2464–9.
27. Koitzsch U, Heydt C, Attig H, et al. Use of the GeneReader NGS System in a clinical pathology laboratory: a comparative study. J Clin Pathol. 2017;70:725–8.
28. Scarpa A, Sikora K, Fassan M, et al. Molecular typing of lung adenocarcinoma on cytological samples using a multigene next generation sequencing panel. PLoS One. 2013;8:e80478.
29. Gargis AS, Kalman L, Bick DP, et al. Good laboratory practice for clinical next-generation sequencing informatics pipelines. Nat Biotechnol. 2015;33:689–93.
30. Roy-Chowdhuri S, Roy S, Monaco SE, et al. Big data from small samples: informatics of next-generation sequencing in cytopathology. Cancer Cytopathol. 2017;125:236–44.
31. Jennings LJ, Arcila ME, Corless C, 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. 2017;19:341–65.
32. Hynes SO, Pang B, James JA, et al. Tissue-based next generation sequencing: application in a universal healthcare system. Br J Cancer. 2017;116:553–60.
33. Malapelle U, Mayo de-Las-Casas C, Rocco D, et al. Development of a gene panel for next-generation sequencing of clinically rel-

evant mutations in cell-free DNA from cancer patients. Br J Cancer. 2017;116:802–10.
34. Fassan M. Molecular diagnostics in pathology: time for a next-generation pathologist? Arch Pathol Lab Med. 2018;142:313–20.
35. Salto-Tellez M. More than a decade of molecular diagnostic cytopathology leading diagnostic and therapeutic decision-making. Arch Pathol Lab Med. 2018;142:443–45.
36. Moore DA, Young CA, Morris HT, et al. Time for change: a new training programme for morpho-molecular pathologists? J Clin Pathol. 2018;71:285–90.