Human tissue in systems medicine

the FEBS Journal
MINIREVIEW

Human tissue in systems medicine

Peter D. Caie[1], Klaas Schuur[1], Anca Oniscu[1], Peter Mullen[2], Paul A. Reynolds[2] and David J. Harrison[1,2]

1 Digital Pathology Unit, Laboratory Medicine, Royal Infirmary of Edinburgh, UK
2 Systems Pathology, School of Medicine, University of St Andrews, UK

Keywords
histopathology; human disease; predictive models; systems biology; systems medicine

Correspondence
D. J. Harrison, Digital Pathology Unit, Laboratory Medicine, Royal Infirmary of Edinburgh, Little France Crescent, Edinburgh EH16 4SA, Scotland, UK
Fax: 00 44 1334 467470
Tel: 00 44 1334 464826
E-mail: david.harrison@st-andrews.ac.uk

(Received 19 April 2013, revised 24 September 2013, accepted 25 September 2013)

doi:10.1111/febs.12550

Histopathology, the examination of an architecturally artefactual, two-dimensional and static image remains a potent tool allowing diagnosis and empirical expectation of prognosis. Considerable optimism exists that the advent of molecular genetic testing and other biomarker strategies will improve or even replace this ancient technology. A number of biomarkers already add considerable value for prediction of whether a treatment will work. In this short review we argue that a systems medicine approach to pathology will not seek to replace traditional pathology, but rather augment it. Systems approaches need to incorporate quantitative morphological, protein, mRNA and DNA data. A significant challenge for clinical implementation of systems pathology is how to optimize information available from tissue, which is frequently sub-optimal in quality and amount, and yet generate useful predictive models that work. The transition of histopathology to systems pathophysiology and the use of multiscale data sets usher in a new era in diagnosis, prognosis and prediction based on the analysis of human tissue.

Introduction

Modern pathology is currently cresting a new technological wave driven by the generation of complex 'big data' through the adoption of novel -omics and digital disciplines. There is a question, however, of how to retrieve optimal useful information from these large data sets in an efficient and clinically relevant manner. The information harvested from the new technologies aligns itself perfectly for the adoption of a systems medicine approach. Although there are differing schools of thought as to how to define systems medicine, in this review we refer to systems medicine as suggested by CASyM: 'Systems Medicine involves the implementation of Systems Biology approaches in medical concepts, research and practice, through iterative and reciprocal feedback between data-driven computational and mathematical models as well as model-driven translational and clinical investigations' (https://www.casym.eu/what-is-systems-medicine). Systems medicine is therefore where specific but large and static data sets acquired across multiple modalities are used to construct computational models for the dynamic prediction of disease progression or response to treatment at a personal level. Systems medicine must be an approach which can be implemented in the clinic and directly benefit patient treatment decisions and outcome. Classical histopathology, however, has in some ways always tried to practice systems medicine. Histopathology, where a pathologist microscopically directly observes the complex diseased tissue system and its interaction with the host microenvironment, and attempts to mentally compute these multiple signals into a prognosis, has long been the gold standard in the clinic. The diagnosis from histopathology has been a model, albeit empirical, seeking to

Abbreviation
EGFR, epidermal growth factor receptor.

derive dynamic meaning of what has happened and what is likely to happen from a static two-dimensional image where secondary information is presented in the artefact that is a tissue section (Fig. 1). From this premise a number of robust, clinically validated tools have evolved with evidence showing that the most simple (e.g. Dukes' staging of colorectal cancer) outlive more complex variants [1]. Standardization of data, and how data are acquired, has long been a goal with internationally agreed criteria for staging cancers such as the TNM classification (tumour size, lymph node status, metastasis) facilitating data transfer between centres and greatly augmenting the possibility of conducting multicentre clinical trials [2]. The advent of the minimal data set brings a deterministic approach to how biopsy tissue samples are reported, and arguably in some situations points the way for non-medical pathologists to do much of the reporting which still remains largely the province of medically trained senior pathologists. In essence this approach is empirical, pragmatic and practical because it allows for the variation that inevitably occurs during tissue accrual and processing, the whole emphasis being what information can be passed back to the clinician. Thus classical histopathology is adept at dealing with imperfection, poor sample, small size of biopsy and poor tissue orientation because the pathologist can easily observe and disregard these artefacts and preen the pertinent information from the stained heterogeneous tissue section. The objective is to frame the tissue along with what it reveals in a useful model which prompts particular actions by the referring clinician.

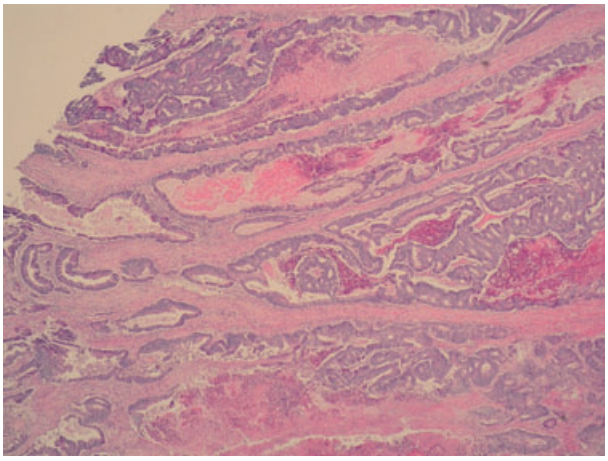

Fig. 1. Standard histological preparation of a colon cancer sample. Irregular glands, necrosis and invasion into the collagenous stroma, combined with a macroscopic assessment of degree of spread, give valuable and accurate prognostic data for categories of patients, but very little personalized information.

However, to train a pathologist may take 5 or more years and is expensive both in training and in maintenance, and the requirement for ever more specialization means that few pathologists now retain a broad perspective on all disease. Instead pathologists are increasingly encouraged to become super-knowledgeable in narrow areas. This brings huge benefits and allows even more accuracy and consistency between centres, but tends to focus more on diagnosis rather than the historical mix of diagnosis plus prediction of what is likely to happen. There are geographical variations in pathology with nomenclature differences between Europe and the USA and different prevalence of some diseases, e.g. in the Far East compared with Europe. But pathologists are a well-connected network and these differences are used constructively to elucidate underlying mechanisms of disease. Similarly, there is growing awareness that animal disease and comparative pathology offer new insights into human pathophysiology [3]. Thus histopathology, as quintessentially a morphological discipline, is entering a new phase with the potential to address phenotype in a clinically useful way whilst also pointing to underlying mechanisms.

Despite the success and evolution of standardized pathology, or perhaps because of it, pathology to some extent has been a bystander for much of the advent of -omics and digital technologies, although it is now joining the movements. Although pharmacogenomics has been around for some time it has had little effect on pathology whereas an explosion in biomarker research is making a definite mark [4]. The current mantra is that a biomarker should accompany a new therapy, suggesting that intrinsically we accept that empiricism is less than adequate for the prescription of new drugs which may be very costly and carry significant side effects. This is an important development and recognizes that histopathology has been very good at making diagnoses, particularly since the advent of routine immunohistochemistry, to elucidate the putative histogenesis of tumour cells and therefore categorize cancer origin and subtype more accurately. This review is not the place to critique whether histogenesis is a meaningful concept in tumour biology, but suffice it to say that increasing knowledge about stem cellness, reprogramming and stem cell niches and their response to treatment, challenges the simplistic assumptions underlying much diagnostic immunohistochemistry-based taxonomy [5]. Standardization of diagnostic labelling and collection of clinical information, in all its forms, means that, in many disease settings, pathology is now also very good at informing prognosis. This is a useful adjunct to planning patient

management but falls some way short of the ideals of 4P medicine, particularly in the areas of predictive, personalized and participatory medicine [6] where a systems medicine approach could prove beneficial. Some immunohistochemical markers such as oestrogen receptor status, in combination with specifically targeted drugs such as oestrogen antagonists, have brought prediction nearer reality; however, the predictive value for an individual is considerably less valuable, in pharmaceutical terms, than the predictive value for a group [7]. But the challenge and promise is clear: pathology can and should inform not just diagnosis and prognosis, but also prediction of what the outcome is going to be after a particular course of treatment. The routine measurement, although we use the term guardedly because it is at best a semi-quantitative ranking algorithm, of oestrogen receptor evolved from more intricate studies of receptor biochemistry [8]. These required fresh tissue, and minimal amounts of cancer versus non-cancerous tissue in the sample, and although the output was a precise estimation the results were grouped to decide treatment. In other words, there were many logistical problems conducting the original biochemical assay which were circumvented by accepting a more robust but much less quantitative assay. This is the crux of understanding the value of histopathology in systems medicine: it is not a question of greater accuracy or precision alone, nor is it the methods to collect tissue in a pristine state, although that is a great goal; rather it is the ability to take whatever information that can be distilled from tissue and use it in a way that wastes nothing, covers doubt and uncertainty and informs clinical management [9]. Pathology is and should be robust in a modelling sense. More recently it has become apparent that the integration of histopathology with other streams of data can significantly add value to, and also challenge and help form, strategies which seek to draw together disparate strands to lead to a diagnosis, inform prognosis and increasingly allow prediction of likely response to therapy. Pathologists already utilize the molecular and morphological heterogeneity in cancer to stratify patients into subgroups with differing prognostic or predictive outcomes. This new pathological knowledge of the difference in a patient's cancer and treatment strategy bears a higher success rate in treating patients and their overall quality of life than a 'one treatment fits all'. This is eloquently exemplified in the treatment of colorectal cancer over the last decade. Previous to targeted biological treatment advanced colorectal patients were treated with FOLFOX or FOLFIRI. Targeted therapy against epidermal growth factor receptor (EGFR) through administration of cetuximab has shown promising results with increased progression-free survival compared with chemotherapy [10]. EGFR inhibition, however, only shows positive results to patients with wild-type KRAS [11,12], and those patients with mutations downstream of EGFR such as KRAS, BRAF [13,14] and PIK3CA [15] show poor response to targeted EGFR monoclonal antibodies such as cetuximab or panitumumab. In fact it has been shown that monoclonal antibody inhibition of EGFR in combination with FOLFOX may actually have detrimental effects on patients with mutant KRAS colorectal cancer [16]. The explosion of data that are being produced and published from proteomic, genomic, transcriptomic and morphometric studies has added value to histopathology. The sheer volume of information, however, may be difficult to handle and real clinical impact may only occur once complex multiple biomarker signatures are combined into a predictive model. It is within these situations that a systems medicine approach may allow multimodality computation of complex data sets and bring clarity to the pathologist and oncologist. There are some key principles learned from pathology that may help facilitate discussion on wider aspects of the application of systems biology approaches in medicine.

The problem of imperfection

Data sets and samples derived from patients are often small or incomplete. The understandable response in seeking to apply systems medicine type approaches is to ensure that more data are collected in a standardized fashion, larger tissue samples are obtained and handled in appropriate ways, and appropriate resource is allocated to the task. As with all automated and computational analysis the quality of the end result is intrinsically linked to the quality of the data input into the model and systems medicine is no different. It makes complete sense to have better data, consistent high quality biological resources and standardization; poor quality data will only result in poor quality outputs. However, the problem with this approach is that ultimately it will fail because some clinical problems are probably not soluble by these means. For example, to investigate signalling pathways it is essential to capture a freeze frame of the precise state of pathway activation, in particular the phosphorylation status of key elements [17]. Much emphasis is given to the retrieval and processing needs of tissue, but almost always this commences after the tissue is removed from the patient and thus surgical ischaemic effects will already have taken place. The resource that has gone into tissue

banking in very many centres is vast [18], but the benefits, although significant, have yet to appear in the clinic and are rather more focused on key centres where tissue collection is immediately linked to tissue use for data generation. The evidence that tissue banking *per se* is beneficial, and a good return on investment, is a moot point. The real problem, especially in large tumours, is that the blood supply to the tissue may be cut off for some minutes before tissue is retrieved, allowing significant transcriptional changes and post-translational modification to occur. So the question arises: how can one use human tissue effectively? The answer depends on what is required. There is no doubt that for discovery, and in the case of systems medicine, optimal tissue and data are required. Tissue banking in such a manner may require large expenditure and a huge logistical effort, but as long as it deals with all the tissue, including the pre-analytical challenges alluded to above, it can result in very high quality data that can be reliably used by the primary investigators but also made available for other groups. For the implementation of systems medicine into routine practice the models and knowledge derived from the gold card approach must be reduced to clinical utility (Fig. 2) [19]. It is unlikely that a patient presenting in the middle of the night with an obstruction in the bowel caused by a cancer will have access to all the facilities necessary for discovery platforms; however, we still want to give that patient the maximum amount of information for their future management. This means that the data collected, even if severely limited, must be fitted to a model and the predictive information derived should be applicable to every patient presenting in clinic. The addition of high throughput technologies such as transcriptomics certainly produces a wealth of data but the sifting of these data, to decide what is needed and what is not, provides additional challenges. So, in many real life settings a systems approach will deliberately limit the amount of data collected or derived from the studied tissue, because too much data introduces the likelihood of error and artefact caused by uncontrollable changes occurring in a sample attained within the clinical setting. To deal with imperfection in clinical samples is thus first to acknowledge it and then to seek to know the correct amount of information, providing that the information obtained is robust and an adequate surrogate, to describe the underlying signalling pathway behaviour. That prompts the question, what is the correct amount of information? This, however, can only be determined empirically and will vary from situation to situation.

Disease is heterogeneous

Histopathology has emphasized the heterogeneity of disease processes, principally by describing differences in morphological appearance. Whilst a primary and metastatic tumour may appear very similar morphologically the molecular phenotype may differ, whereas there may also be marked variation in appearance within a single primary tumour [20,21]. But pathology does not necessarily reflect underlying genetic heterogeneity, and genetic heterogeneity may not be reflected in differing morphology [22]. Furthermore, heterogeneity may exist at many different levels due to mutation status, stem cell niche, hypoxic areas, chromatin structure or methylation, and the science of relating these levels of data to the objectives of 4P medicine are largely unexplored. Some models have extrapolated from a cell-based biochemical signalling pathway to a cancer tissue without perhaps accounting in full for the heterotypic nature of cancer (composed of cancerous and non-cancerous elements) and its heterogeneity

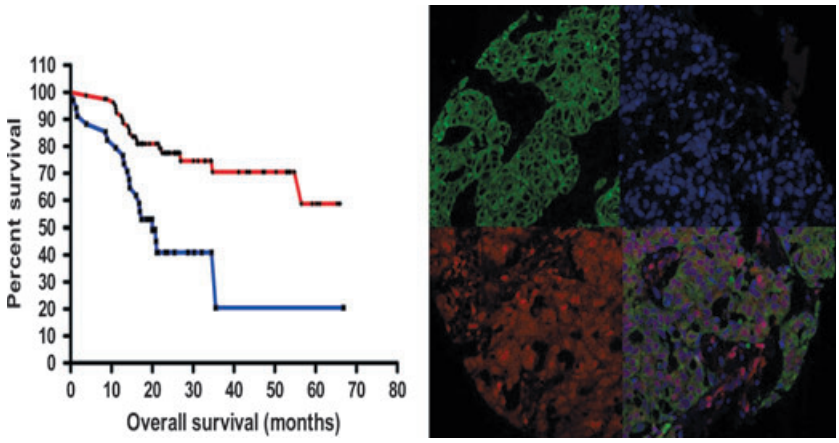

Fig. 2. Clinical impact on the stratification of patients from quantified immunofluorescence protein expression. The left panel shows the Kaplan–Meier survival curve for breast cancers treated with herceptin segregated according to PTEN expression. The right panel shows quantitative, multichannel fluorescence with DAPI stained nuclei in blue, tumour cells in green and PTEN expression in red. (Figure reproduced from [11].)

(variation in genetic aberration across different parts of the tumour). Histopathology has itself not adequately resolved how heterogeneity affects outcome (prognosis or prediction) despite years of empiricism, in part due to problems of lack of standardization and consistency. This is an exciting challenge for systems medicine on a multiscale level. There are very practical issues to consider. For example, if a mutation in EGFR is crucial to determine whether a patient should receive a drug to treat lung cancer [23] then one has to question whether it is legitimate to assume that a tiny biopsy, perhaps < 100 cells, is sufficient to categorize a tumour as mutant or not. The imperfection of clinical practice is that no more tissue is available and so a huge assumption has to be made. The corollary of this is that too large a sample of tumour may give evidence of heterogeneity that we are uncertain how to use. Interestingly pathology has already solved this problem empirically as its scoring systems for predictive biomarkers usually combine an indication of both intensity and variability of expression [6]. Whether this empirical solution, which works for groups of patients, can be reduced to work at a personalized level needs to be investigated. This discussion is largely relevant to cancer tissues but the same general principles apply to any study relying on tissue, with organs such as brain showing marked variation in structure, appearance and function in health and disease. A further important consideration is what part of a tumour is actually important. In many studies, certainly in those requiring transcript analysis from homogenized samples, there is an assumption that the mass effect should mean something. However, if < 1% of a tumour is composed of the cells that actually generate the progeny that cause growth, spread and ultimately death we may be spending inordinate amounts of money to get the right answer to the wrong question. Laser capture microdissection of heterogeneous subpopulations within the tissue section allows its purification and quantification for predictive studies. Similarly whole slide imaging coupled with automated image analysis and spatial statistics delivers continuous immunohistochemical and morphological data from heterogeneous tissue and information on how it interacts with itself and the host. These techniques may yield a higher chance of success for systems medicine as heterogeneity will affect the entire model and its ability to predict the long-term effectiveness of set treatments for overall tumour reduction and long-term repression at the personal level. Systems medicine must go hand in hand with a greater understanding of tumour biology in all its variation and complexity to adopt a truly systems approach.

Pathology is a two-dimensional artefact

Histopathology for much of its existence as a discipline has relied upon a two-dimensional artefact, stained with antiquated histochemical techniques and viewed by transmission light microscopy to deduce behaviour, both past and future, in a dynamic three-dimensional way. The addition of monoclonal antibodies and latterly RNA and DNA extraction have allowed in-depth analyses of some components, but bizarrely histopathology for all of this still remains the best working model available for many diseases. It is apparent that the artefact of a tissue section is in fact a distillation of a complex phenotype, a surrogate of DNA, RNA, proteome, methylome, metabolome all rolled into one without any clear guidebook of how to balance different features. The application of quantitative automated image analysis techniques, e.g. morphometry and quantitative immunofluorescence, encourages the pathologist to provide more dynamic raw data to the modeller allowing greater awareness of the spatial and presumably temporal variation that occurs [8]. Automated image analysis of cellular phenotypes, or high content biology, has already been adopted and implemented within pharmaceutical drug discovery pipelines [24]. The application of digital and quantitative techniques to pathology has long been heralded but is slow to arrive in routine use [25]; however, we are now seeing the rapid emergence of the new field of digital pathology. In part this may be due to the problems discussed above where the object has been to achieve perfection of data input rather than utility of model output. A recent study implemented unsupervised automated image analysis and discovered that the stromal features, i.e. non-tumour cell characteristics, of breast cancer showed a marked relationship to prognosis [26]. Whilst there is a risk in these studies of overfitting data to solutions, this shows an exciting step forward in image analysis. Previous studies going back decades have associated nuclear size with outcome in diseases as disparate as bladder and ovarian cancer [27,28]. Measuring nuclear DNA ploidy using flow cytometry was a hugely popular pastime for a while but delivered almost nothing of generalizable clinical value [29]. Pathologists of course had got there first, with the ordinal scale of tumour grade as an estimate of nuclear size and presumably aggressiveness! In many situations this is indeed the case, but not always, perhaps pointing to differences in underlying mechanisms. More recently digital image analysis and parallel computing have allowed greater study of quantitative characteristics of cancer [30]. Whilst

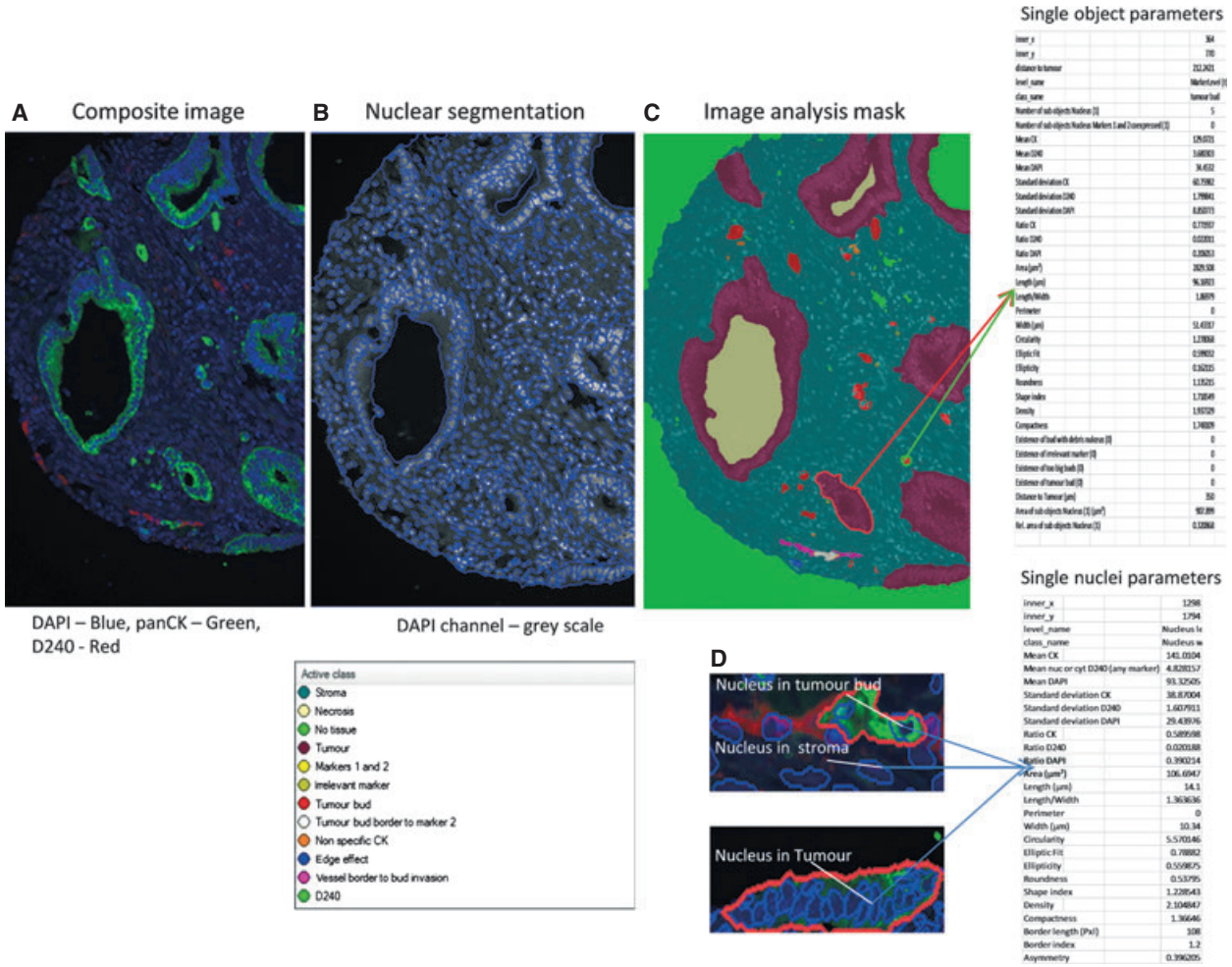

Fig. 3. Automated quantification of prognostic histopathological features in a colorectal tissue microarray (TMA) core through digital pathology. The TMA was stained by immunofluorescence using antibodies against pan-cytokeratin (epithelium, green) and D240 (lymphatic endothelium, red) and counterstained with DAPI (nucleus, blue). (A) The image was captured and digitized prior to importing into Definiens Developer XD™ software. (B) Utilizing the DAPI image layer all nuclei within the TMA were segmented and quantified for intensity and morphometric parameters. (C) Combining the three image layer wavelengths the image was first segmented into stroma, tumour and necrosis/lumen. Second, tumour buds and lymphatic vessels were quantified through the pan-cytokeratin (marker 1) and D240 (marker 2) image layers respectively. A combination of epithelium, as tumour and tumour bud, and D240 colocalization (markers 1 and 2) along with quantification and classification of bordering epithelial and D240 markers was employed to quantify LVI (lymphatic vessel invasion). Intensity, spatial and morphometric single object parameters may be quantified from within the tumour and its microenvironment. (D) To quantify tumour grade and the nuclei within the invasive microenvironment while retaining their heterogeneity, single object nuclei intensity, texture and morphometric parameters were quantified and classified within separate subpopulations of 'tumour', 'stroma' and 'tumour bud'.

unsupervised studies have their place there is also a move to using image analysis to identify and measure features of interest such as lymphovascular invasion, tumour budding and nuclear grade in colon cancer (Fig. 3). Whether these measured features mean anything, in the sense that they contribute to refining the personalized and predictive nature of the diagnosis, remains to be seen but the advent of these approaches is perfect for systems medicine as they encourage questions, experimentation and iteration of a model and thus fulfil classical criteria for the application of systems approaches. Greater automation is unlikely to replace pathology for some time if at all, but the incorporation of fully quantifiable and continuous data, not otherwise captured in an appropriate model, may further harness the value of tissue morphology which despite all the advances in molecular biology still remains the best single investigative tool we have. All measurements and assessments made on tissue are based on an artefact, a surrogate of what is really

happening, fixed at one point in time. Using that knowledge opens new conceptual avenues.

Pathology is imaginary!

But sometimes dreams come true. Histopathology is essentially visual, confirming that meaning inferred may be more rich and complex than simply multiple measured data sets. Thus, the future of systems pathology and medicine may be to visualize processes, building in methods that allow filling-in where data are incomplete and prompting focus of research efforts where gaps appear in the picture. Systems approaches are showing us how we can use complex large data sets to build useful models to better understand disease and how to treat it. Clinical implementation will require a sifting of those data to ensure that (a) models work and can be iteratively improved and refined, (b) data that are quality assured and sufficient can be gathered in real clinical settings and (c) we keep an awareness that we have an empirical tool in pathology that works remarkably well and that we should seek to add value to it rather than simply to substitute an alternative.

References

1 Logan RF, Patnick J, Nickerson C, Coleman L, Rutter MD & von Wagner C; English Bowel Cancer Screening Evaluation Committee (2012) Outcomes of the Bowel Cancer Screening Programme (BCSP) in England after the first 1 million tests. *Gut* **61**, 1439–1446.

2 Sobin LH, Gospodarowicz MK & Wittekind C (2009) Classification of Malignant Tumours. 7th edn. Wiley-Blackwell, Oxford.

3 Calvisi DF, Frau M, Tomasi ML, Feo F & Pascale RM (2012) Deregulation of signalling pathways in prognostic subtypes of hepatocellular carcinoma: novel insights from interspecies comparison. *Biochim Biophys Acta* **1826**, 215–237.

4 http://www.cancerresearchuk.org/prod_consump/groups/cr_common/@fre/@fun/documents/generalcontent/cr_027486.pdf.

5 Sadanandam A, Lyssiotis CA, Homicsko K, Collisson EA, Gibb WJ, Wullschleger S, Ostos LC, Lannon WA, Grotzinger C, Del Rio M *et al.* (2013) A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat Med* **19**, 619–625.

6 Hood L & Flores M (2012) A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory. *N Biotechnol* **29**, 613–624.

7 Molina R, Ciocca DR, Tandon AK, Allred DC, Clark GM, Chamness GC, Gullick WJ & McGuire WL (1992) Expression of HER-2/neu oncoprotein in human breast cancer: a comparison of immunohistochemical and western blot techniques. *Anticancer Res* **12**, 1965–1971.

8 Hawkins RA, Hill A & Freedman B (1975) A simple method for the determination of oestrogen receptor concentrations in breast tumours and other tissues. *Clin Chim Acta* **64**, 203–210.

9 Faratian D, Clyde RG, Crawford JW & Harrison DJ (2009) Systems pathology – taking molecular pathology into a new dimension. *Nat Rev Clin Oncol* **6**, 455–464.

10 Jonker DJ, O'Callaghan CJ, Karapetis CS, Zalcberg JR, Tu D, Au HJ, Berry SR, Krahn M, Price T, Simes RJ *et al.* (2007) Cetuximab for the treatment of colorectal cancer. *N Engl J Med* **357**, 2040–2048.

11 Karapetis CS, Khambata-Ford S, Jonker DJ, O'Callaghan CJ, Tu D, Tebbutt NC, Simes RJ, Chalchal H, Shapiro JD, Robitaille S *et al.* (2008) K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med* **359**, 1757–1765.

12 Amado RG, Wolf M, Peeters M, Van Cutsem E, Siena S, Freeman DJ, Juan T, Sikorski R, Suggs S, Radinsky R *et al.* (2008) Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J Clin Oncol* **26**, 1626–1634.

13 Loupakis F, Ruzzo A, Cremolini C, Vincenzi B, Salvatore L, Santini D, Masi G, Stasi I, Canestrari E, Rulli E *et al.* (2009) KRAS codon 61, 146 and BRAF mutations predict resistance to cetuximab plus irinotecan in KRAS codon 12 and 13 wild-type metastatic colorectal cancer. *Br J Cancer* **101**, 715–721.

14 Di Nicolantonio F, Martini M, Molinari F, Sartore-Bianchi A, Arena S, Saletti P, De Dosso S, Mazzucchelli L, Frattini M, Siena S *et al.* (2008) Wild-type BRAF is required for response to panitumumab or cetuximab in metastatic colorectal cancer. *J Clin Oncol* **26**, 5705–5712.

15 Sartore-Bianchi A, Martini M, Molinari F, Veronese S, Nichelatti M, Artale S, Di Nicolantonio F, Saletti P, De Dosso S, Mazzucchelli L *et al.* (2009) PIK3CA mutations in colorectal cancer are associated with clinical resistance to EGFR-targeted monoclonal antibodies. *Cancer Res* **69**, 1851–1857.

16 Douillard JY, Siena S, Cassidy J, Tabernero J, Burkes R, Barugel M, Humblet Y, Bodoky G, Cunningham D, Jassem J *et al.* (2010) Randomized, phase III trial of panitumumab with infusional fluorouracil, leucovorin, and oxaliplatin (FOLFOX4) versus FOLFOX4 alone as first-line treatment in patients with previously untreated metastatic colorectal cancer: the PRIME study. *J Clin Oncol* **28**, 4697–4705.

17 Pinhel IF, Macneill FA, Hills MJ, Salter J, Detre S, A'hern R, Nerurkar A, Osin P, Smith IE & Dowsett M

(2010) Extreme loss of immunoreactive p-Akt and p-Erk1/2 during routine fixation of primary breast cancer. *Breast Cancer Res* **12**, R76.
18 Speirs V & Morgan A (2013) Breast cancer: investment biobanking – increased returns from tissue samples. *Nat Rev Clin Oncol* **10**, 128–129.
19 Faratian D, Goltsov A, Lebedeva G, Sorokin A, Moodie S, Mullen P, Kay C, Um IH, Langdon S, Goryanin I *et al.* (2009) Systems biology reveals new strategies for personalizing cancer medicine and confirms the role of PTEN in resistance to trastuzumab. *Cancer Res* **69**, 6713–6720.
20 Aitken SJ, Thomas JS, Langdon SP, Harrison DJ & Faratian D (2010) Quantitative analysis of changes in ER, PR and HER2 expression in primary breast cancer and paired nodal metastases. *Ann Oncol* **21**, 1254–1261.
21 Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, Kamiyama M, Hruban RH, Eshleman JR, Nowak MA *et al.* (2010) Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467**, 1114–1117.
22 Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P *et al.* (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366**, 883–892. erratum **367**, 976.
23 Lindeman NI, Cagle PT, Beasley MB, Chitale DA, Dacic S, Giaccone G, Jenkins RB, Kwiatkowski DJ, Saldivar JS, Squire J *et al.* (2013) 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* **8**, 823–859.
24 Caie PD, Walls RE, Ingleston-Orme A, Daya S, Houslay T, Eagle R, Roberts ME & Carragher NO (2010) High-content phenotypic profiling of drug response signatures across distinct cancer cells. *Mol Cancer Ther* **9**, 1913–1926.
25 Ghaznavi F, Evans A, Madabhushi A & Feldman M (2013) Digital imaging in pathology: whole-slide imaging and beyond. *Annu Rev Pathol* **8**, 331–359.
26 Yuan Y, Failmezger H, Rueda OM, Ali HR, Gräf S, Chin SF, Schwarz RF, Curtis C, Dunning MJ, Bardwell H *et al.* (2012) Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Sci Transl Med* **4**, 157ra143.
27 Brinkhuis M, Scheepstra C, Buist MR, van Diest PJ & Baak JP (1997) Intratumor heterogeneity of morphometric and stereologic variables in primary ovarian tumors and their omental metastatic deposits. *Anal Quant Cytol Histol* **19**, 185–193.
28 Kapur U, Antic T, Venkataraman G, Durazo-Arvizu R, Quek MM, Flanigan RC & Wojcik EM (2007) Validation of World Health Organization/International Society of Urologic Pathology 2004 classification schema for bladder urothelial carcinomas using quantitative nuclear morphometry: identification of predictive features using bootstrap method. *Urology* **70**, 1028–1033.
29 Laerum OD & Farsund T (1981) Clinical application of flow cytometry: a review. *Cytometry* **2**, 1–13.
30 Lloyd MC, Allam-Nandyala P, Purohit CN, Burke N, Coppola D & Bui MM (2010) Using image analysis as a tool for assessment of prognostic and predictive biomarkers for breast cancer: how reliable is it? *J Pathol Inform* **1**, 29.