Next Generation Pathology

Peter D Caie, David J Harrison

Systems Pathology

School of Medicine

University of ST Andrews

North Haugh

St Andrews

Fife, UK

KY16 9TF

Summary

The field of pathology is rapidly transforming from a semi-quantitative and empirical science toward a Big-Data discipline. Large data-sets from across multiple –omics fields may now be extracted from a patient's tissue sample. Tissue is, however, complex, heterogeneous and prone to artefact. A reductionist view of tissue and disease progression, which does not take this complexity into account, may lead to single biomarkers failing in clinical trials. The integration of standardised multi-omics Big-Data and the retention of valuable information on spatial heterogeneity is imperative to model complex disease mechanisms. Mathematical modelling through systems pathology approaches is the ideal medium to distil the significant information from these large, multi-parametric and hierarchical data-sets. Systems pathology may also predict the dynamical response of disease progression or response to therapy regimens from a static tissue sample. Next generation pathology will incorporate Big-Data with systems medicine in order to personalise clinical practise for both prognostic and predictive patient care.

Key words: histopathology, integrative pathology, systems pathology, spatial heterogeneity, predictive models, cancer pathology, multi-omics, image analysis

1. Introduction

The manual, microscopic viewing of thinly cut and stained tissue sections by histopathologists has been the steadfast method of deciphering tissue architecture and concluding a prognosis for multiple diseases for over one hundred years. The field of pathology recognises the rich data source which lies within a tissue section. With the aid of specific histochemical stains, augmented by immunological and even mRNA or DNA based approaches, the pathologist takes into account the entire heterogeneous and heterotypic microenvironment and its interactions across the tissue section. Through experience they are able to process this complex, sometimes subtle information, and translate it in order to aid their diagnostic or prognostic conclusion. Research pathologists also apply this methodology to evaluate novel or significant prognostic features such as the tumour differentiation, tumour gland morphology at the invasive front or immune infiltrate within the microenvironment. The development of immunohistochemistry from the 1940s provided the pathologist with the ability to interrogate the tissue section with a further level of complexity where they could match biomarker expression with histopathological features and morphometry, although it took some time and the advent of monoclonal antibodies some 30 years or so later for a dramatic increase in routine use of the technology. The use of protein biomarkers, visualised through immunohistochemistry, allowed quantification at both spatial heterogeneity and subcellular resolution. Since the post-omics era the field of modern pathology is experiencing an explosion of data across multiple but disparate –omics strands. Most notably within the clinic is the genomic profiling of a patient's tissue sample through next generation sequencing (NGS) where, in colorectal cancer for example, EGFR and KRAS mutations now may be routinely tested for in order to predict the response to anti-EGFR antibody treatment. Single "magic-bullet" biomarkers, however, have a limited use in clinical prognosis, drug prediction and efficacy studies as they attempt to describe or modulate complex multi-pathway molecular and cellular interactions in an often too simplistic way.

Advances in the integration of genomics, proteomics, transcriptomics, epigenomics and the emerging field of image analysis based phenomics are now able to add valuable information to the hierarchical

understanding of complex disease mechanisms. These molecular signatures correlated with morphological and clinical data have the ability to advance traditional diagnostic medicine from broad population-based prediction to a more personalised and precision based science. Pathology has overcome the bottleneck of creating large, hierarchical and complex “Big-Data”, however the challenge the field is now facing is how to handle this data in a meaningful manner which directly leads to translational impact. The over-arching goal of modern big-data pathology is to infer a dynamical prediction of disease from a static patient tissue sample. Systems pathology through mathematical modelling allows the integration, interrogation and identification of significant parameters from large multi-omics data sets while having the ability to add a dynamic aspect to personalised medicine.

2. Tissue is Heterogeneous

Tissue is extremely heterogeneous and cancer especially so; cancer heterogeneity can originate from multiple sources: cell of origin, clonal evolution, cancer stem cells (CSC), response to microenvironment and host factors as well as stromal or immune cell infiltrate. The clonal evolution theory states that the cancers build up heterogeneous subpopulations after concurrent mutations over multiple rounds of cell division due to the plasticity of the cells through chromosomal and replicative instability or exogenous insults. These heterogeneous subpopulations are under the influence of natural selection where they may acquire mutations which ultimately lead to cell death while others accumulate a specific set of driver mutations allowing the cancer cells to metastasise. CSCs may originate from healthy tissue stem cells or may have attained their stem-like phenotype through epigenetic alterations of the genome or through stromal cell interaction from their microenvironmental niche. The stem-like attributes associated with CSCs would confer a certain amount of plasticity upon it in order for it to evade aggressive treatment regimens or commit to the metastatic cascade. CSCs may have the ability to produce hierarchical heterogeneous cell subpopulation progenies of which only some are tumourigenic and others differentiated. CSCs are thought to initiate tumourigenesis, have the ability to propagate the cancer after chemotherapy and a cure for the patient depends on the eradication of such self-renewing cells. CSCs also appear to be

more resistant to radiation and chemotherapeutic treatment and may incur tumour recurrence even after a long period of remission and dormancy. More recently the “Big Bang model” of intra-tumour heterogeneity has been described where tumours mainly grow as a single expansion and that intra-tumour heterogeneity within tumour subpopulations is high but occurs early on in the tumour’s evolution. In this model aggressive subclones may not be predominant and can remain undetected although they would provide overall resistance to subsequent insult by treatment regimens (1).

The focus of cancer research for prognosis, prediction and drug discovery has been on the tumour itself however this target is changing. It is becoming apparent that the tumour microenvironment as a whole, and more precisely the stromal and immune infiltrate, is increasingly important in tumour progression and evasion of chemotherapy. The host interaction on the tumour, their stem-cell subpopulations and their microenvironmental niche adds a further level of heterogeneity to the tumour. Spatial heterogeneity within the stromal compartment of the tumour is a critical influence on the tumour, its subsequent progression and potential resistance to therapy. The combination of the above creates a further level of complexity in the accurate understanding of disease and for its dynamic modelling.

3. Tissue samples are imperfect

A wealth of prognostic and predictive information lies within the patient’s tissue sample. Classical histopathology strives to infer dynamical prediction of disease progression from the static artefact which is the tissue section. The pathologist directly observes microscopically the complex diseased tissue and its interaction with the host microenvironment in order to mentally compute these multiple signals into a prognosis. This has long been the gold standard in clinical prognosis. Although multiple novel prognostic methodologies for Colorectal Cancer (CRC) have been developed to replace or augment classical pathology, and while some show promise, for example the gene expression signatures ColoPrint (2) and Oncotype DX (3), none has established itself within routine clinical prognosis. The classical Dukes and TNM morphological and histological staging of the disease remains steadfast in clinical pathology. One reason for this is standardisation and the imperfection of

tissue. The human eye can account for the variation and artefacts that occur between surgical removal of the tissue through to mounting sections onto microscope slides for analysis. Poor and small sample size, imperfection and damage to tissue as well as poor tissue orientation can be easily disregarded by the pathologist while they can glean the pertinent information from the final stained tissue section. Automated quantification of the tissue section, spanning the -omics fields, is not able to be so selective and may therefore return variable results. The need for standardisation across all aspects of automated tissue datafication is therefore essential.

Advances in extracting data in a meaningful and robust manner will add value to classical histopathology methodologies and provide greater impact and accuracy of patient stratification at a more personalized level than current population statistics, such as TNM staging. This is increasingly relevant when the quantification techniques take into account the heterogeneity of the disease and report on it. Datafication of tissue is the extraction of information in a fully quantifiable and standardised manner. This can take the form of quantifying a single biomarker to capturing a complex and hierarchical multi-modal omics signature. Routinely, single read-outs are extracted from a single tissue sample however advances in data capturing technologies now allow multiple readouts captured across multiple –omics fields which may be reported across distinct subpopulations identified through morphometric or biomarker expression. Big-data pathology is now a reality but creating standardized data sets amenable to complex modelling and which take into account the imperfection of tissue and its inherent heterogeneity is still in its infancy.

4. Quantifying heterogeneity

Understanding tumour heterogeneity is important in striving toward an intelligent and individualised treatment strategy which translates into clinical impact. To truly fulfil a personalised medicine approach and select the correct combination therapy for a patient it is essential to know which mutational or epigenetic aberrations their cancer carries in both primary and distant disease and what the subsequent phenotypic and functional effect on the cells and their microenvironment are.

Multiple interactions at multiple levels occur in tissue architecture. Histopathology describes the end result but not the underlying molecular mechanisms. Since the post ‘omics’ era scientists have been armed with a suite of new tools to identify biomarkers to subgroup a patient’s cancer at the molecular level. Using these tools a raft of data and new biomarkers have been discovered over the last few decades and allowed genome scale analysis and comparisons. The main disciplines to bear the wealth of the results are genomics, transcriptomics, proteomics and epi-genetics. Technologies such as NGS and arrayCGH allow the mutation and copy number status of the genome to be analysed. RNA microarray chips and RNA-sequencing technologies are employed to profile gene expression whereas Reverse phase protein array (RPPA) and mass spectrometry have brought proteomics into the field of Big-Data pathology.

Inter-patient and intra-patient heterogeneity exists (Figure 1) and the aim of all ‘omics’ research is to identify biomarkers which can lead to targeted drug discovery programmes or companion diagnostics which will allow the clinician and pathologist to make rapid informed decisions on the prognosis of the disease and to predict which treatment will display the greatest efficacy and best outcome as possible for the individual patient.

Although the above methodologies to quantify the molecular mode of action driving cancer subtypes have added significant value, they also hold disadvantages to assaying such complex material. To extract DNA, RNA and protein molecules these assays usually homogenise and destroy the tissue integrity. The tissue is literally “mashed and measured” mixing together any subpopulations of cancer and host cells expressing differential properties while losing spatial resolution. This results in one end-point being reported for the whole tumour. Due to the nature of these applications, intra-tumoural heterogeneity of the tissue may be under-detected where the dominant or most abundant genotypes or phenotypes mask signal from smaller cell populations within the tumour. Healthy tissue and host cells from the tumour microenvironment are both also added to the molecular sample creating a further source of noise to the signal and could increase the reporting of false positive or negative results. Under-detection of tissue heterogeneity therefore leads to an urgent and difficult problem when treating a patient with combination therapy, as resistant subgroups could go unnoticed and untreated.

There are, however, tools to overcome this problem which attempt to better quantify, and thus comprehend, the complexity of heterogeneous tumours. One such tool is laser capture microdissection (LCM) which isolates and analyses cells and sections of the tissue of interest, usually those displaying morphological differences. This technique allows the separate analysis of distinct subpopulations as well as comparing the tumour's core, invasive edge and the stromal microenvironment. From these distinct sections DNA, RNA and protein can be isolated and studied resulting in a cleaner profile of the difference between regions of interest and their heterogeneity. Recent technological advances allow molecular genomic and proteomic profiling with small sample sizes amenable to ever smaller tissue samples which is advantageous in the study of heterogeneity. Background signals from complex tissue can still create noise in these assays and robust and sensitive data depends on the LCM technique as well as the specificity of probes, antibodies and detection technology used. To avoid contamination of signals, from heterogeneous subpopulations within tissue, *in situ* imaging of protein through Immunohistochemistry (IHC) and genomics through fluorescence *in situ* hybridisation (FISH), may be applied. This has advantages over destructive assays as the tissue structure, spatial orientation and sub-localisation of molecules are retained and heterogeneity can be visualised, compartmentalised and quantified while providing insight into cellular interactions within the tumour and its microenvironment. IHC further allows the visualisation of morphological status of the cells expressing the biomarker of interest and allows the observer to correlate morphometric and proteomic signatures at the cellular resolution. Spatial heterogeneity impacts the prognostic and predictive significance of biomarkers and it is becoming increasingly apparent that this must be taken into consideration for the modelling of disease. The immunoscore in colorectal cancer, which quantifies the density and intra-tumoural location of CD3+ and CD8+ lymphocytes through image analysis, has been shown to hold a higher prognostic significance than the gold standard of TNM staging (4).Similarly, the spatial heterogeneity of unbiased and automatically quantified lymphocytes in breast cancer tissue sections was statistically modelled and found to be associated to patient survival(5). In the field of transcriptomics it has recently been discovered that mesenchymal cell gene expression classifiers are linked to poor prognosis in colorectal cancer though it proves difficult to ascertain whether these classifiers are expressed by the tumour or the stromal cells however

immunohistochemistry for mesenchymal proteins in tissue sections as well as laser capture microdissection have elucidated that the mesenchymal signatures originate from stromal cancer-associated fibroblasts and not from the tumour itself (6, 7).

Although the field of high content analysis is not new, where multiple parameters and biomarkers are measured from fluorescently labelled cells(8), the discipline has been slow to translate to histopathology and the clinic. This has been in part due to the complexity of tissue and its imperfection compared to *in vitro* cell studies and the need for extensive validation and standardisation for clinical use. This is now changing and digital pathology as well as automated image analysis for tissue-based studies is rapidly emerging into the realm of clinical research. The integration of digital pathology with automated image analysis brings advantages to the field. These include the standardisation of quantification where observer variability is excluded and the robust analysis of rare or complex features is captured. Traditionally image analysis in histopathology concentrated on the quantification of protein expression through immunohistochemistry and immunofluorescence (IF). This was to overcome the subjective manual and semi-quantitative scoring of a 1+, 2+, 3+ system. Upon employing IF, image analysis software can perform fully quantified continuous data-exports from which cut-offs can be calculated in order to stratify patient subgroups. Computer based quantification of nuclear morphometry, however, has been practiced for over a decade. Continuous improvements to image analysis software now allow the simultaneous export of morphometric parameters of cells and histopathological features alongside biomarker quantification associated to this feature. In this co-registering methodology it is possible to identify surrogate morphological features which correlate with molecular phenotype.

The market leading tissue imaging platform manufacturers provide their own image analysis solutions for chromogenic and fluorescence assays which allows segmentation of cells and subcellular compartments and subsequent biomarker quantification within heterogeneous tissue. These software packages are designed to work in connection with the images captured from their own platforms and can sometimes be restrictive to the quantification of set assays, biomarkers and parameters. Definiens (http://www.definiens.com/), Indica lab (http://indicalab.com/) and Visiopharm

(http://www.visiopharm.com/) offer image analysis packages which can import images from most microscopes and allow a more flexible image analysis environment to capture the complexity of the heterogeneous tissue section.

While the imaging of a single biomarker can yield predictive or prognostic information the ability to multi-plex two or more markers on a single tissue section becomes a much more powerful tool. An advantage of IF based image analysis is the ability to multiplex, co-register and quantify biomarkers at the cellular resolution. Multi-plexing reports on protein interactions, pathway activation and multiple cellular events. Accurate co-localisation and spatial resolution of multiple biomarkers or histological features on the same section of tissue reports a richer high content and functional data than serial sections of one biomarker while saving the precious resource which is the tissue sample. Researchers can quantify multiple proteins on a per cell basis or accurately quantify multiple cell types within a heterogeneous population. Traditional multiplexing is limited by bleed through of fluorophores and chromagens as well as antibody cross-reactivity of secondary host-species. Multi-spectral imaging and un-mixing of chromagens and fluorophores allows an accurate spectral readout for each biomarker of interest, increases the multi-plexing capacity and negates any autofluorescence. Sophisticated image analysis software and multi-plexed *in situ* labelling permit the big-data capture from image analysis based segmented tissue sections to quantify the data-rich histopathology and the interactions and spatial heterogeneity of the cancer microenvironment's phenotypic features. This involves the extraction of complex and hierarchical data pertaining to a single segmented feature or set of features across the segmented tissue section. This data may be captured through co-registering of biomarkers as proteomic or genomic signals, as multiple morphometric and texture parameters or a combination of both; essentially extracting as much data as possible from each single segmented object within the image. A multi-parametric signature is therefore built up for each tissue sample which may be compiled of multi-omic image based features. Tissue subpopulations may be identified in this manner and further mined through *in situ* labelling or microdissection to interrogate the patient's sample at the personalised level for predictive or prognostic pathology (Figure 2).

Sophisticated data mining is required to identify the significant single or combination of parameters within the signature in order to stratify patients for prognostic or predictive purposes. Data mining techniques previously applied to identify significant parameters have been logistic regression analysis and ensemble decision tree models. Further advancements in *in* situ labelling and image analysis such as mass spectrometry imaging, Next-generation immunohistochemistry (9) and multi-parametric data capture, where biomarkers are correlated to morphometry, are catapulting this field into the realm of true Big-Data alongside the more traditional –omics fields. Image analysis and *in situ* labelling of tissue sections coupled to spatial statistics will most probably factor highly when profiling a diseases complex heterogeneous microenvironment in the future of systems pathology.

5. Integrative pathology

Traditional omics research attempts to identify single molecular or histopathological features which could be utilized for prognosis or prediction of response to drug therapy. Cancer is, however, a very complex disease with multiple molecular interactions within the cell and multiple cellular interactions within the microenvironment. Many single biomarkers never translate to the clinic, as they do not take into account the complexity and heterogeneity of the disease. Integrating large scale data from multiple omics fields may help to address this problem as it will create a better understanding of the multiple molecular interactions occurring within the cell and how these translate to disease progression. This approach was exemplified in colorectal cancer where histopathological subtypes were integrated with methylation and mutation status to assess their correlation and impact on prognosis (10). Integrative large scale pathology has also been implemented in breast cancer where cellular resolution of *in situ* and co-registered genotype and phenotype was utilised to study intra-tumoural heterogeneity between primary and distant metastasis for studies of prognosis and potential drug targets (11). Finally, a further breast cancer study integrated a multi-omics signature and discovered JAK-STAT and TNF signalling pathways to be significant in triple negative disease which could lead to novel and personalised drug treatments (12). There is a wealth of data collected during classical histopathology which largely remains unused in clinical decision making. This clinical data

is beginning to be integrated with the modern datafication modalities as a further hierarchical level of understanding of the disease from the tissue. In mucoepidermoid carcinoma; histopathology, immunophenotypic and cytogenetic parameters were integrated to identify a signature which was able to identify the pulmonary disease from other subtypes of lung cancer (13). Clinical and molecular data is now also being integrated with the complex and data-rich image-based phenotypic signatures to investigate cancer heterogeneity and its interaction with the microenvironment. The morphometric signatures can also be correlated to the genomic profile and clinical outcome (5). Computational IT solutions are also now available which allow the incorporation of multiscale omics data (14, 15) as well as integrate it with clinical information (16).

6. Systems pathology

Pathology is now adept at creating large and complex data sources from across the omics fields and more recently including histopathology, morphometrics and spatial heterogeneity. This data, however, must be integrated in a meaningful way which makes best use of its complexity, is standardized, reproducible and robust enough to be clinically relevant. The challenge ahead is how to incorporate this integrated data into models which can identify the optimal combinations of parameters to answer clinical questions in a robust and standardised manner. Systems medicine, and more recently systems pathology, takes a holistic view of tissue, the cell and its multitude of interactions. Systems pathology requires a large amount of high-quality multi-scale data to be extracted from tissue and which acts as input for predictive mathematical models. Although systems pathology has predominantly concentrated on molecular profiling of the genome, transcriptome or proteome, image analysis based multi-parametric biomarker and morphometry is perfectly matched to add to the hierarchical data within a systems model. This additional *in situ* information allows the retention of the valuable spatial heterogeneity within the diseases microenvironment.

Essentially, a modern integrative pathology would adopt the principles of 4P medicine in a systems pathology approach. 4P medicine consists of Prediction, Personalisation, Prevention and Patient participation (17). There are many definitions of systems medicine. Within Europe systems medicine

is defined by the EU consortium CASyM (www.casym.eu), as stated within the first chapter of this book.

The principle of systems pathology is to predict a dynamic pathological response from static data sets. The more standardised and robust the data which is used for input into the model directly relates to the quality of prediction within the model. Systems pathology is complex with the implementation of multiple differential equations into a multiscale dynamic model to predict a drug effect on a patient or inform how that patient will respond over time. Systems pathology, under this definition, was utilised to confirm the role of PTEN in Trastuzumab drug resistance (18). Systems pathology can also be implemented to track tumour evolution post chemotherapy through intra-tumour heterogeneity and spatial distribution of phenotype and genotype at the cellular level (19). In CRC a systems pathology approach was employed to identify a disease recurrence signature in early stage patients from a multi-omics data set where parameters associated with immune response were found to be the most significant predictors (20).

Systems pathology is therefore already making a valuable impact into the field of translatable clinical research. Systems pathology is the ideal tool to distill significant parameters with significant population cut-offs, and which are therefore translatable to the clinic, from multiple integrated complex Big-Data sets. This is what we have termed ‘Next Generation Pathology’. The ultimate goal of next generation pathology is to make use of this hierarchical data captured across multiple modalities from an imperfect and static tissue sample, in order to better understand both disease progression and a patient’s personalised response to treatment.

References

1. Sottoriva A, Kang H, Ma Z, Graham TA. A Big Bang model of human colorectal tumor growth. 2015;47(3):209-16.
2. Kopetz S, Tabernero J, Rosenberg R, Jiang ZQ, Moreno V, Bachleitner-Hofmann T, et al. Genomic Classifier ColoPrint Predicts Recurrence in Stage II Colorectal Cancer Patients More Accurately Than Clinical Factors. Oncologist. 2015;20(2):127-33.
3. Srivastava G, Renfro LA, Behrens RJ, Lopatin M, Chao C, Soori GS, et al. Prospective multicenter study of the impact of oncotype DX colon cancer assay results on treatment recommendations in stage II colon cancer patients. Oncologist. 2014;19(5):492-7.
4. Galon J, Mlecnik B, Bindea G, Angell HK, Berger A, Lagorce C, et al. Towards the introduction of the Immunoscore in the classification of malignant tumors. J Pathol. 2013.
5. Yuan Y. Modelling the spatial heterogeneity and molecular correlates of lymphocytic infiltration in triple-negative breast cancer. J R Soc Interface. 2015;12(103).
6. Isella C, Terrasi A, Bellomo SE, Petti C, Galatola G, Muratore A, et al. Stromal contribution to the colorectal cancer transcriptome. 2015.
7. Calon A, Lonardo E, Berenguer-Llergo A, Espinet E, Hernando-Momblona X, Iglesias M, et al. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. 2015.
8. Caie PD, Walls RE, Ingleston-Orme A, Daya S, Houslay T, Eagle R, et al. High-content phenotypic profiling of drug response signatures across distinct cancer cells. Mol Cancer Ther. 2010;9(6):1913-26.
9. Rimm DL. Next-gen immunohistochemistry. Nat Methods. 2014;11(4):381-3.
10. Inamura K, Yamauchi M, Nishihara R, Kim SA, Mima K, Sukawa Y, et al. Prognostic significance and molecular features of signet-ring cell and mucinous components in colorectal carcinoma. Ann Surg Oncol. 2015;22(4):1226-35.
11. Almendro V, Kim HJ, Cheng YK, Gonen M, Itzkovitz S, Argani P, et al. Genetic and phenotypic diversity in breast tumor metastases. Cancer Res. 2014;74(5):1338-48.
12. Karagoz K, Sinha R, Arga KY. Triple negative breast cancer: a multi-omics network discovery strategy for candidate targets and driving pathways. Omics. 2015;19(2):115-30.
13. Roden AC, Garcia JJ, Wehrs RN, Colby TV, Khoor A, Leslie KO, et al. Histopathologic, immunophenotypic and cytogenetic features of pulmonary mucoepidermoid carcinoma. Mod Pathol. 2014;27(11):1479-88.
14. Le Cao KA, Gonzalez I, Dejean S. integrOmics: an R package to unravel relationships between two omics datasets. Bioinformatics. 2009;25(21):2855-6.
15. Day RS, McDade KK, Chandran UR, Lisovich A, Conrads TP, Hood BL, et al. Identifier mapping performance for integrating transcriptomics and proteomics experimental results. BMC Bioinformatics. 2011;12:213.
16. Miyoshi NS, Pinheiro DG, Silva WA, Jr., Felipe JC. Computational framework to support integration of biomolecular and clinical data within a translational approach. BMC Bioinformatics. 2013;14:180.
17. Hood L, Friend SH. Predictive, personalized, preventive, participatory (P4) cancer medicine. Nat Rev Clin Oncol. 2011;8(3):184-7.
18. Faratian D, Goltsov A, Lebedeva G, Sorokin A, Moodie S, Mullen P, et al. Systems biology reveals new strategies for personalizing cancer medicine and confirms the role of PTEN in resistance to trastuzumab. Cancer Res. 2009;69(16):6713-20.
19. Almendro V, Cheng YK, Randles A, Itzkovitz S, Marusyk A, Ametller E, et al. Inference of tumor evolution during chemotherapy by computational modeling and in situ analysis of genetic and phenotypic cellular diversity. Cell Rep. 2014;6(3):514-27.

20. Madhavan S, Gusev Y, Natarajan TG, Song L, Bhuvaneshwar K, Gauba R, et al. Genome-wide multi-omics profiling of colorectal cancer identifies immune determinants strongly associated with relapse. Front Genet. 2013;4:236.

Figure captions

Figure 1. Inter- and Intra-patient heterogeneity

A) Loss of E-Cadherin at the invasive front of CRC. A TMA core taken from the invasive front of a CRC patient tumour block. Neoplastic glands are visualized with antibody against panCK (green) and counterstained with DAPI (blue). E-Cadherin (red) is not expressed in neoplastic glands at the edge of the cancer invasion (green box) but is expressed in well differentiated glands located closer to the tumour centre (red box).

B) Cytokeratin 7 inter-patient heterogeneity. TMA cores taken from 2 different patient blocks: core A and core B. Core B shows high expression Cytokeratin 7 (red) in the neoplastic cells (green) and Core A shows no Cytokeratin 7 expression.

Figure 2. Subpopulation segmentation and biomarker quantification through image analysis

Tumour subpopulation segmentation and classification through whole slide image analysis of immunofluorescence labelled colorectal cancer tissue utilizing Definiens image analysis software. A) raw image: DAPI (blue) and panCK (green). B) Image analysis algorithm automatically segments tumour from stroma. C) Tissue is further segmented into stromal cells, tumour buds, poorly differentiated clusters (PDC) and three tumour gland subpopulations. Ki67 (red) proliferation marker is quantified within separate subpopulations at the invasive front (F & G) and the the tumour core (D & E).