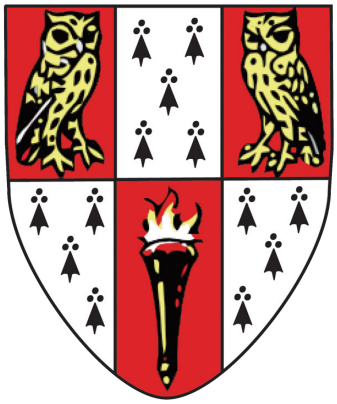
CANCER CARTOGRAPHY

Mapping Cancer Evolution in Tissue Space



Artem Lomakin

European Bioinformatics Institute, EMBL-EBI
Hughes Hall, University of Cambridge

This thesis is submitted for the degree of
Doctor of Philosophy

September 2023

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is the result of my own work and any work done in collaboration is declared at the beginning of each chapter. This thesis does not exceed the prescribed word limit of 60,000 words excluding tables, footnotes, bibliography, and appendices as defined by the Degree Committee for the Faculty of Biology.

Abstract

Cancer Cartography: Mapping Cancer Evolution in Tissue Space
Artem Lomakin

While somatic evolution is widely accepted as fundamental to cancer development, significant gaps in understanding persist. These gaps concern the role of cellular interactions within the tumour microenvironment (TME) and the impact of spatial constraints on cancer evolution. Emerging spatial omics technologies offer the potential to address these gaps, although their application in spatial genomics remains limited. This is particularly crucial because genetic alterations not only drive cancer evolution but also serve as an archive of its history.

To address this research gap, this thesis aims to develop computational methods for analysing spatial genomics data, specifically on data generated by base-specific *in situ* sequencing (BaSISS), a method that enables high-resolution mapping of diverse somatic mutations across large tumour tissue sections. Bayesian algorithms designed in the study generate quantitative clonal maps, effectively tracing cancer evolution in tissue space while accommodating multiple forms of biological and technical variability. Despite inherent assumptions, the algorithm demonstrates robustness and quantitative accuracy.

Complementary data types contribute to the findings by applying the Bayesian model to two multifocal breast cancers at different stages of progression. Integration of histology, immunohistochemistry (IHC), and targeted *in situ* gene expression allows for the phenotypic characterisation of clones in distinct microanatomical niches. Subsequent analyses across various stages of breast cancer, including carcinoma *in situ*, invasive cancer, and lymph node metastasis reveal clone-specific variations in proliferation, morphology, stroma, hypoxia, and immune microenvironments. In one instance involving ductal carcinoma *in situ*, polyclonal neoplastic expansions manifest on a macroscopic scale but remain segregated within microanatomical structures.

In summary, this thesis establishes a robust computational framework for extracting clonal architecture from spatial genomics data. It provides a proof-of-concept that such maps, when integrated with tissue morphology and spatial phenotype data, can offer vital insights into the mechanisms driving both cancer evolution and tissue ecology.

Acknowledgements

I could not have accomplished the work presented in this thesis without the exceptionally collaborative, intellectually stimulating, and supportive environment I have been privileged to experience throughout my PhD journey. First of all, I would like to thank my supervisors and "scientific parents" Moritz Gerstung and Lucy R. Yates. Your balanced mentorship – allowing autonomy while offering guidance – has been invaluable. I also thank the Gerstung group for their *esprit de corps* and engaging lunch and coffee-break conversations. I thank my collaborators Artem and Gleb from Dissatisfaction.ai team, and Mikhail S. Batov. Our side projects offered a much-needed break from the primary research. I must also thank Vitalii for our stimulating dialogues on mathematical models, biology and life.

I appreciate the sense of community and social engagement within the EBI predoc group. This has created a relaxed social atmosphere which made my PhD time much more enjoyable. I would also like to thank the EBI cycling club members – Lukas, Leah, Paula, and Felix – for the enjoyable rides and slipstreams during our commutes.

Finally, my deepest gratitude goes to Paula for her unwavering love and support, and to my family for staying close despite being so far away.

Contents

List of Figures

List of Abbreviations

ADVI automatic differentiation variational inference

BaSISS base-specific *in situ* sequencing

BFB breakage-fusion-bridge

CAF cancer-associated fibroblast

CCF cancer cell fraction

CIS carcinoma *in situ*

CNA copy number alteration

DCIS ductal carcinoma *in situ*

DP Dirichlet process

DTC disseminated tumour cell

ECM extracellular matrix

ELBO evidence lower bound

EMT epithelial-mesenchymal transition

FDR false discovery rate

GLMM generalised linear mixed model

GP Gaussian process

H&E haematoxylin and eosin

HLA human leukocyte antigen

HMC Hamiltonian Monte Carlo

IHC immunohistochemistry

ISS *in situ* sequencing

LCM laser capture microdissection

LN lymph node

MCMC Marcov chain Monte Carlo

PBC primary invasive breast cancer

PVL perivascular-like cell

scRNA-seq single-cell RNA sequencing

SNV single nucleotide variant

TDLU terminal ductal lobular unit

TIL tumour infiltrating lymphocyte

TLS tertiary lymphoid structure

TME tumour microenvironment

TNBC triple-negative breast cancer

VAF variant allele frequency

VI variational inference

WGS whole genome sequencing

Spatial Biology of Cancer Evolution: A review

1

Contributions

This chapter draws upon the work, for which I am the co-first author contributing equally with Zaira Seferbekova, detailed in:

Zaira Seferbekova, Artem Lomakin, Lucy R Yates and Moritz Gerstung [May 2023]. 'Spatial biology of cancer evolution'. en. In: *Nat. Rev. Genet.* 24.5, pp. 295–313.

I conducted literature research in collaboration with Z.S., and we both wrote the manuscript under the supervision of L.R.Y. and M.Ger.. Based on a discussion among all authors, Z.S. created the initial figures, which a professional illustrator later redrew.

The content of this chapter has been adapted from the original review manuscript, with only minor stylistic and structural alterations. The concluding Section 1.7 is original.

1.1. Components of cancer evolution

The drivers of evolution are *mutation* and *selection* [J. Cairns 1975; Nowell 1976]. Mutation of somatic cells is an inevitable and persistent consequence of life [R. Li et al. 2021; Martincorena and Campbell 2015; Moore et al. 2021]. In most normal tissues, mutations accumulate at a steady rate of around 15–50 per cell per year of life, with only the germline known to exhibit lower rates [Moore et al. 2021]. Tumour cells often exhibit an elevated mutation burden due to the variety of mutational processes they have been exposed to during life and, in some cases, due to acquired hypermutation [Alexandrov et al. 2020]. The continuous

mutation *Changes in the DNA sequence that are inherited across cell generations. Sources of mutation are erroneous replication, biochemical alterations of DNA and failed DNA repair. To date, all cells within the human body are found to accumulate mutations over their lifetime.*

selection *In evolution, natural selection denotes the process of survival and reproduction of the fittest organism within a given environment.*

accumulation of mutations inevitably leads to diversification at the level of single cells both in tumours and normal tissues.

The second force of evolution is selection, which describes how a fitter lineage outgrows its relatives. Selection operates at the level of a wide range of heritable phenotypes that derive from genomics and epigenomics. Clones with a selective advantage in the environment to which they are exposed will expand, while those with a disadvantage will tend to disappear (Figure 1.1). Advantageous variants are therefore enriched in genomes of aged normal tissues and cancers [Greenman et al. 2006; Martincorena et al. 2017]. More than 500 so-called cancer driver genes have been reported to date [Gonzalez-Perez et al. 2013; Lawrence et al. 2013; Sondka et al. 2018], which are believed to cause different cancer hallmark traits and enabling characteristics that involve cell-intrinsic mechanisms as well as interactions with the TME [Hanahan et al. 2000; Douglas Hanahan 2022].

Clones *A clone is a population of cells that derive from a common ancestor. Cancers are found to be clones deriving from a single cell that expanded during the lifetime of a host. Clonal alterations are mutations that are present in all cells of a cancer because their occurrence preceded the expansion of the tumour.*

TME *A combination of non-tumour cells, such as stromal and immune cells, vessels, metabolites, signalling molecules, and other extracellular components among which tumour cells exist.*

subclones *Further clones emerging within a tumour from one founder clone. Subclonal mutations are limited to a fraction of cancer cells and occur during tumour expansion.*

positive selection *The spread of advantageous alleles within a population.*

While mutation continuously generates subclonal diversity at the level of single cells, it is well documented that tumours are mosaics of subclones each comprising hundreds of thousands of cells that have arisen from a shared ancestor [Andor et al. 2016; Dentro et al. 2021; McGranahan et al. 2015; Nik-Zainal, Van Loo et al. 2012; Shah et al. 2009] (Figure 1.1); these patterns have been further confirmed by single-cell studies [Casasent et al. 2018; Laks et al. 2019; McPherson et al. 2016; Navin et al. 2011; Salehi et al. 2021; Y. Wang et al. 2014]. The mechanisms of these expansions remain debated. It has been shown that cancer subclones exhibit signs of positive selection and driver gene mutations that can also be found clonally [Dentro et al. 2021]. However, it is also conceivable that subclones branching at early stages of tumour development reach considerable size without a selective advantage as they continue to expand with the same tumour, a phenomenon termed neutral evolution [M. J. Williams et al. 2016].

Subclonal mosaicism is often found to be spatially variegated as demonstrated in diverse studies based on tumour macrodissection [Bruin et al. 2014; Gerlinger et al. 2014; 2012; Jamal-Hanjani et al. 2017; Morrissy et al. 2017; Navin et al. 2010; Watkins et al. 2020; Yates et al. 2015] and laser-capture microdissection [Bao et al. 2018; Casasent et al. 2018; Grossmann et al. 2021; Heide et al. 2022; 2019; F. Su et al. 2018; Woodcock et al. 2020; Zhao et al. 2022] (Figure 1.1). These clonal variegation patterns were shown to be accompanied by differential gene expression in multiple cancer types, including renal [Gerlinger et al. 2012], colorectal [Househam et al. 2022], lung [Biswas et al. 2019] and breast cancer [Lomakin et al. 2022], using spatially resolved genomic and transcriptomic analyses, indicating the presence of subclone-specific gene expression and associations between

certain subclones and characteristic TMEs.

The exact mechanisms creating these phenomena are not fully understood, largely owing to a dearth of methods that can molecularly characterize whole tumour sections with spatial resolution. A variety of non-exclusive explanations have been proposed. A rapid expansion could lead to a fragmentation of closely related clones [Sottoriva et al. 2015] even though it has also been argued that cell dispersal is a crucial element shaping tumour structure [Gallaher et al. 2019; Waclaw et al. 2015]. However, it is also possible that tissue micro-anatomy, such as the ductal system of the breast, provides a template for clonal segregation (discussed further in the Section 1.2). Lastly, it is conceivable that locally distinct microenvironments or niches favour the selection of particular clones.

Various new spatially resolved genomic, transcriptomic and proteomic technologies offer novel insights and answers to these unresolved questions (Figure 1.1). The technological aspects enabling the nascent field of spatial cancer biology are summarised in Box 1 and have been extensively reviewed elsewhere [Lewis et al. 2021]. Individual technologies differ in the multiplexity of their readouts and their spatial resolution as well as in their sensitivity and available field of view. Combining genomic, transcriptomic and proteomic layers enables a rich characterisation of tumour ecosystems.

tumour ecosystems *The collective set of heterogeneous cells in the vicinity of a tumour comprising cancer cells and the TME.*

This introduction focuses on the spatial aspects of cancer evolution, summarises how spatial transcriptomics and proteomics reveal a complex landscape of the tumour ecosystem, discusses how interactions with the TME are integral to cancer evolution, and proposes potential clinical applications and future directions. Particular focus is given to the role of tissue micro-anatomy in controlling the rate of evolution, as well as cellular interactions with the TME within which cancer cells evolve.

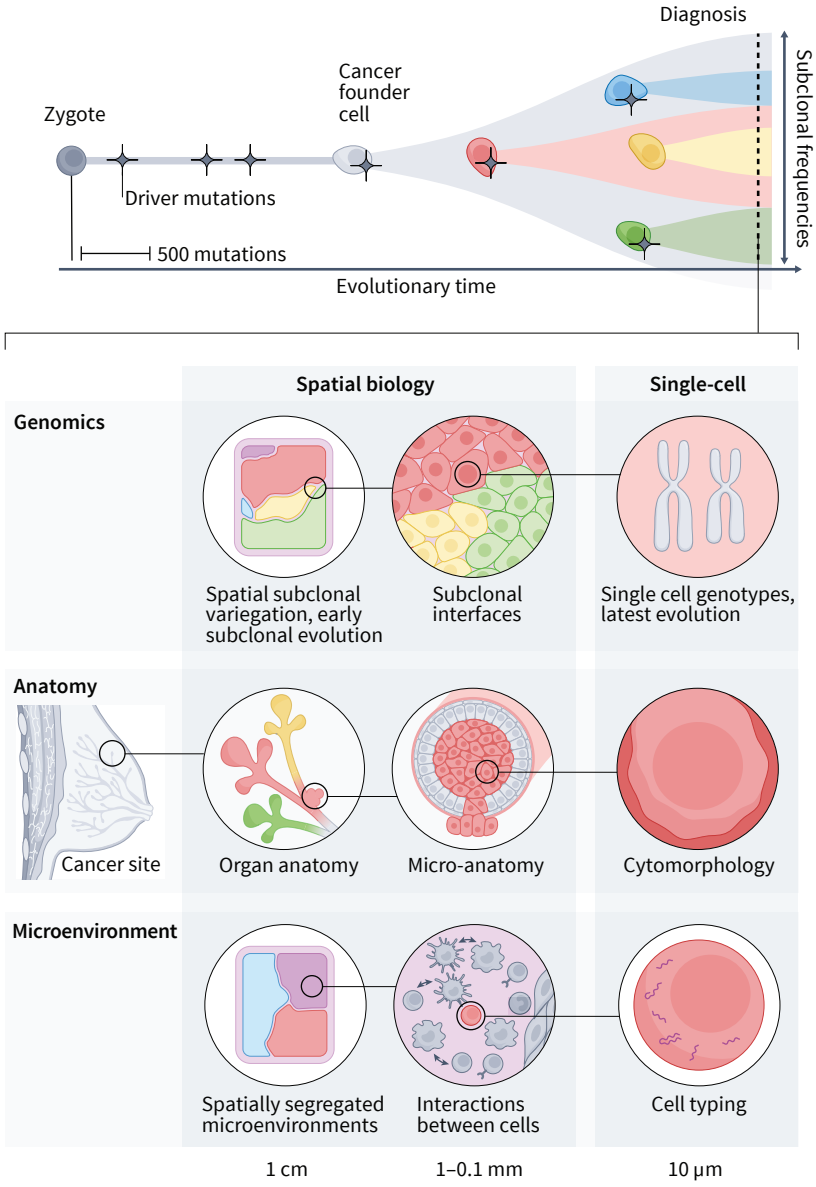


Figure 1.1: Spatial biology of cancer evolution [Seferbekova et al. 2023]. ***top***, Somatic evolution and cancer development. Human cells accumulate mutations throughout their lifetime. A subset of these, so-called driver mutations, are associated with neoplastic transformation. Ongoing mutation and selection cause subclonal divergence within the tumour clone. ***bottom***, Spatial technologies add new layers of information. A tumour is a mosaic of subclones with distinct micro-anatomy and tumour microenvironment. Spatial analyses offer new avenues for charting and understanding the biology and evolution of the cancer ecosystem at scales ranging from a single cell to whole tissue.

1.2. Tissue organisation controls evolution

The rate at which malignant clones emerge and spread through the tissue is controlled by its organisation. Levels of organisation include the micro-anatomical tissue architecture as well as differentiation hierarchies with few stem cells feeding successively larger pools of differentiated cells. The breakdown of these protective principles is a key feature of cancer development. However, the resident tissue structure, for example, the ductal system in the breast or epithelial layers of the oesophagus, can also influence the rate of progression at pre-invasive stages. The transformation ends with overwhelming metastatic disease.

tissue architecture *The micro-anatomical spatial organization of the tissue. Typical examples are layered epithelial tissues, glands and crypts. Tissue architecture can be combined with other means of tissue organization such as differentiation hierarchies.*

Normal tissue organisation suppresses evolution

It has long been hypothesized that somatic tissues are structured to suppress the rate of somatic evolution [J. Cairns 1975]. These ideas are based on fundamental theoretical considerations that the rate of evolution depends on population structure. Generally, the rate of evolution is determined by the rate at which new mutations are generated (a product of the number of cells and the mutation rate per cell), the probability that new mutants sweep through the population [Lieberman et al. 2005] and the time it takes to do so [Frean et al. 2013; Tkadlec et al. 2019]. In addition to suppressing the mutation rate, differentiation hierarchies can reduce the chance of emergence and slow the spread of mutant lineages. The haematopoietic system constitutes a basic example [Lopes et al. 2007]: its 10^{13} cells derive from around 100,000 haematopoietic stem cells [Lee-Six et al. 2018; Sender et al. 2016], which themselves divide slowly and produce a range of faster-dividing differentiated cells [Humphries et al. 2008]. This hierarchy reduces the rate of evolution, as only mutations in the comparably small number of stem cells and potentially the first progenitors, or mutations that lead to a differentiation blockage, will remain in the population. The colon extends these organisational principles by its compartmentalization into micro-anatomical crypts; in each crypt, a similar differentiation hierarchy operates in which a very small number of stem cells replenishes the colonic epithelium [Humphries et al. 2008]. As differentiating cells are fated to die within 5 days, only mutations arising in the stem cells persist [J. Cairns 1975; Nowak et al. 2003; Sender et al. 2021]. Furthermore, micro-anatomy largely constrains the fixation of stem-cell mutations to individual crypts (Figure 1.2).

Systematic studies of normal tissues confirm that, in tissues with a more complex architecture, clones typically remain small in size and develop independently

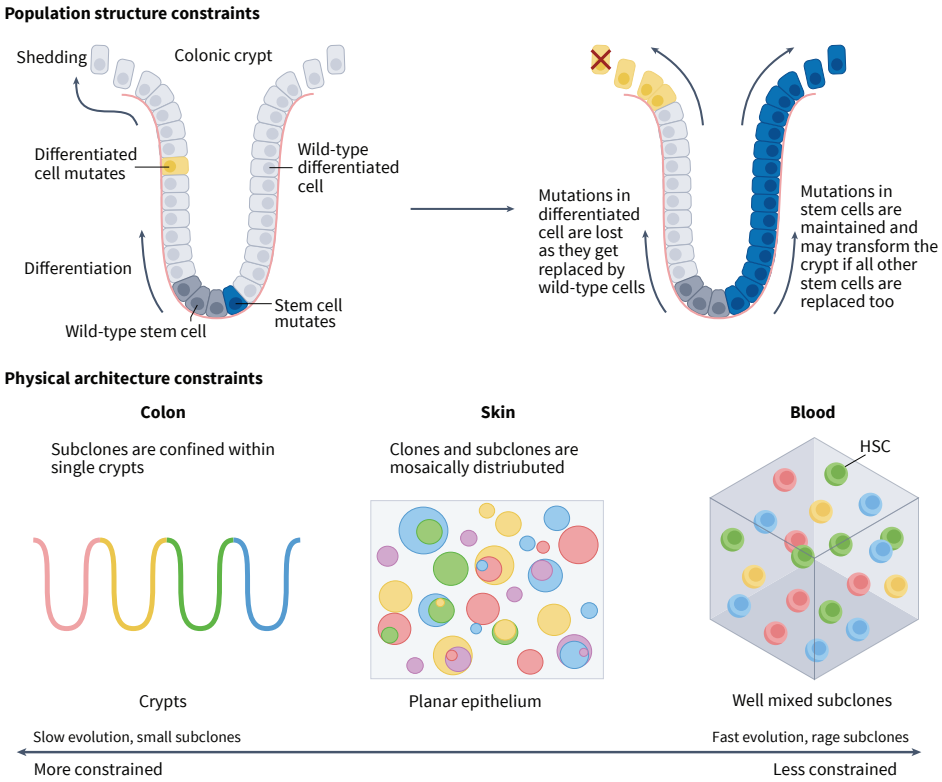


Figure 1.2: Normal tissue structure suppresses somatic evolution [Seferbekova et al. 2023]. ***top***, In addition to physical constraints, the differentiation hierarchy in the colonic crypt limits the number of cells with fixation potential, which reduces the chances of malignant transformation. ***bottom***, Highly structured tissues, such as the colon, slow evolution by constraining the spread of clones. Physically unstructured tissues, such as the haematopoietic system, exhibit the fastest rate of evolution and clones, which replace the entire organ

in comparison to less structured tissues where clonal expansions tend to be larger [R. Li et al. 2021]. In a recent investigation of 517 normal colonic crypts, only 1% harboured cancer driver gene mutations, and those that were mutated were confined to single crypts [Lee-Six et al. 2019]. Similar observations have been made for glandular tissues such as the endometrium [Moore et al. 2020]: although most of the glands harboured driver mutations, they were limited to a single mutation or a few neighbouring mutations. Conversely, it is well established that the haematopoietic system harbours mutant lineages whose frequencies can reach up to 100% [Genovese et al. 2014; Jaiswal et al. 2014]. Other epithelial tissues, including the skin, oesophagus or bladder, have an intermediate degree of tissue

structure and are patchworks of microscopic clones [Lawson et al. 2020; Martincorena et al. 2018; Martincorena, Roshan et al. 2015; Moore et al. 2021] (Figure 1.2). In such tissues, the fixation probability is high but the time to fixation is greater than in the haematopoietic system because the dynamics are confined to two dimensions, which reduces the interfaces at which a fitter variant may replace less-fit competitors. Of note, not all mutations causing a selective advantage are associated with malignant transformation. For example, in the skin and oesophagus, *NOTCH1* mutant clones expand and suppress malignant evolution while maintaining normal tissue function [Abby et al. 2021; Colom et al. 2021; Fowler et al. 2021; Martincorena et al. 2018; Martincorena, Roshan et al. 2015].

Breakdown of evolutionarily confining tissue architectures in cancer

The fundamental observation that dedifferentiation and the loss of normal tissue architectures are correlated with cancer aggressiveness was made a century ago and still forms part of modern histopathological grading systems that direct cancer treatment [Elston et al. 1991; Epstein et al. 2016; Greenough 1925; Louis et al. 2007]. In these systems, higher grades correspond to more aggressive tumours, typically composed of less differentiated cells.

The breakdown of such evolutionarily confining tissue structures is a hallmark of cancer. It is especially noticeable among carcinomas, which derive from epithelial tissues (Figure 1.3). At the earliest stages of progression, so-called carcinoma *in situ* are still restrained to the resident tissue structures, such as the duct of the breast. Similar to the considerations of normal tissues, the resident tissue structures may delay clonal sweeps. This tissue-based restraint was observed in a case of breast cancer, where spatial genomic analysis revealed two distinct subclones occupying a chequerboard pattern of micro-anatomical niches within the ductal system [Lomakin et al. 2022].

carcinoma in situ *Neoplastic expansions of epithelial cells that are confined to the normal tissue structure in which they arose, without invasion of the adjacent stroma.*

During invasion, tissue architectures are lost, probably via different mechanisms in each tissue type, but a frequent property of carcinomas is the epithelial-mesenchymal transition (EMT). EMT enables epithelial cells to lose their polarity and escape from the confinement of planar epithelial tissue organisation (as reviewed in Polyak et al. [2009]). In colorectal adenocarcinoma, EMT can be induced via the *WNT* signalling pathway by loss-of-function mutations in *APC* or somatic mutations of *CTNNB1* inhibiting its degradation [Morin et al. 1997]. Both alterations are believed to be a key step in the development of adenomas, the neoplastic precursor to invasive adenocarcinoma. A further step of invasion is the breakdown and remodelling of the extracellular matrix (ECM), thought to also be

invasion *The process of penetration and spread of cancer cells into the neighbouring normal tissue. In the case of carcinomas, invasion involves a breach of the basement membrane.*

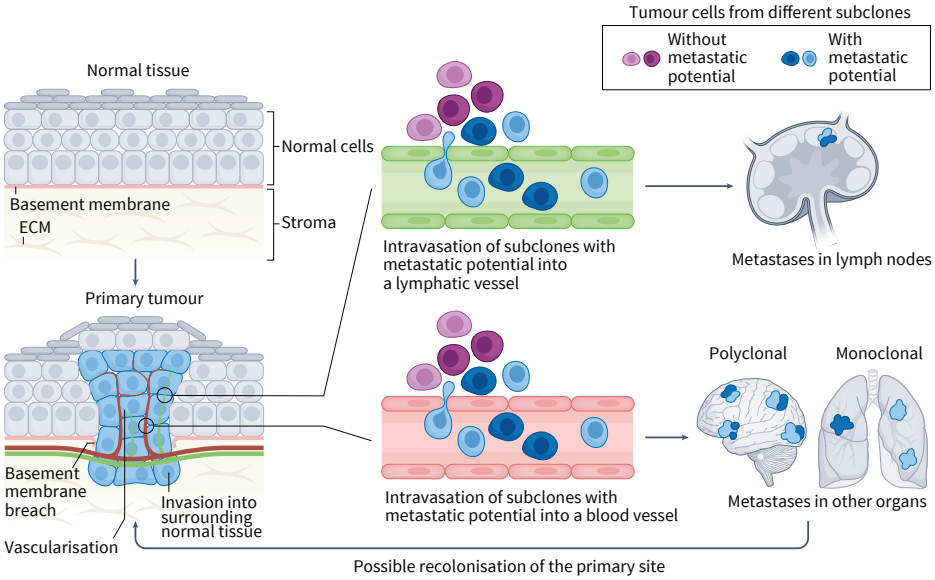


Figure 1.3: Spatial steps of cancer progression [Seferbekova et al. 2023]. Normal tissue structure breaks down during tumour progression. This process is exemplified here by a carcinoma: breach of the basement membrane is followed by tumour invasion into the neighbouring stroma, lymphangiogenesis and angiogenesis. Through blood and lymphatic vessels, cancer cells with metastatic potential can transfer to lymph nodes and distant organs, giving rise to metastases at these new sites. Metastases containing cancer cells of different lineages within the same lesion are called polyclonal, whereas metastases containing cancer cells of the same lineage are called monoclonal. Metastatic subclones may later return to the primary site and recolonize it. ECM, extracellular matrix; HSC, haematopoietic stem cell

facilitated by the microenvironment (see the Section 1.3).

The consequence of this breakdown is likely the acceleration of malignant evolution as the fixation probability and the rate of fixation are higher in unstructured populations with a greater degree of mixing. Several theoretical investigations have demonstrated that spatial constraints can control the mode of tumour evolution [Noble et al. 2022; West et al. 2021]. Another consequence is that different selective pressures apply, which may be one cause of the characteristic sequences of cancer progression such as the classic *APC-KRAS-TP53* model for colorectal carcinoma [Fearon et al. 1990; Gerstung et al. 2020] and the four stages of *TP53* inactivation in the progression of pancreatic ductal adenocarcinoma from a pre-malignant to malignant state [Baslan et al. 2022]. In this light, a pan-cancer analysis revealed that the portfolio of late and subclonal driver gene mutations tends to be more diverse than the set of early alterations [Gerstung et al. 2020].

Metastatic dissemination

The ability of cancer cells to leave a primary tumour site, seed in distant organs and generate metastatic deposits presents landmark stages that are strongly predictive of poor clinical outcomes [Lambert et al. 2017]. This inherently spatial process involves, upon tissue invasion, intravasation of single or small aggregates of cancer cells, circulation through the blood system, arrest in the capillaries of the target organ, extravasation, and proliferation [Fidler 2003] (Figure 1.3). These steps thus involve a range of micro-anatomical bottlenecks and possible differential selective pressures in different stages and target organs.

Several multi-region genomic sequencing studies have investigated the phylogenetic relationships of primary tumours and metastases from the same patient [Bruin et al. 2014; Gerlinger et al. 2012; Jamal-Hanjani et al. 2017; Jones et al. 2008; Karlsson et al. 2018; Mamlouk et al. 2017; Noorani et al. 2020; Yates et al. 2015; 2017; Huaqiang Zhou et al. 2021]. These data show that most metastases have diverged from the primary tumour clone typically shortly before or after its most recent common ancestor, which is often dated 1–5 years before diagnosis [Gundem et al. 2015; Hu et al. 2020; Noorani et al. 2020; Yates et al. 2017]. The origin of metastasis appeared to be mostly clonal [Brastianos et al. 2015; Brown et al. 2017; De Mattos-Arruda et al. 2019] even though, in advanced disease, intermixed subclones have been found in multiple metastases, suggesting spread from one metastasis to another [Aceto et al. 2014; Gundem et al. 2015; Noorani et al. 2020]. Pan-cancer analyses of metastatic cancers have revealed that metastases exhibit fewer intra-metastatic subclones than primary tumours [Martínez-Jiménez et al. 2022; B. Nguyen et al. 2022; Priestley et al. 2019], which may in part be a consequence of sampling fewer cells (often core needle biopsies). From the perspective of driver gene mutations, it has been argued that treatment-naive metastasis has broadly the same driver gene mutations as the matched primary [Reiter et al. 2019]. Characteristics of metastatic cancer include high levels of genomic instability, *TP53* mutations and whole-genome duplications, shared by both the primary tumour and metastases [B. Nguyen et al. 2022; Priestley et al. 2019]. However, adjuvant therapies can select for specific mutations, for example, prostate and breast cancers treated with endocrine deprivation therapies are enriched for resistance mutations in *AR* and *ESR1*, respectively [Martínez-Jiménez et al. 2022; B. Nguyen et al. 2022].

most recent common ancestor *In cancer evolution, the most recent founder cell from which all other tumour cells have directly descended.*

metastasis *A process of cancer spreading from the primary disease site to lymph nodes or other organs, resulting in the formation of metastatic deposits known as metastases.*

adjuvant therapies *Cancer treatments given after primary therapy (particularly following surgical resection of a tumour) to reduce the risk of relapse.*

Although it emerges that tissue architecture has a profound influence on somatic evolution, it is important to realise that cellular interactions with normal cells are also likely to affect the fate of cancer clones. The composition of the

TME and its role in cancer evolution will be discussed in the next two sections.

1.3. Mapping the cancer ecosystem

The cellular composition of tumour masses is highly variable, with the neoplastic lineage contributing as few as 10% of cells in some cases [Aran et al. 2015]. The remaining cells derive from the TME, which is usually dominated by immune and stromal cells in addition to vessels, adipocytes and pericytes. Together, these cells and other physical structures, such as ECM and chemical gradients, form an environment termed the cancer ecosystem [Somarelli 2021]. Over the past decade, single-cell RNA sequencing (scRNA-seq) analyses have catalogued the cellular composition of the cancer ecosystem [Chung et al. 2017; Darmanis et al. 2017; Izar et al. 2020; Lambrechts et al. 2018; Pombo Antunes et al. 2021; Puram et al. 2017; Tirosh et al. 2016; Venteicher et al. 2017]. More recently, spatial transcriptomic and proteomic technologies have started to uncover how these diverse cell types organise into communication interfaces and niches [Arnol et al. 2019; Schapiro et al. 2017] (Boxes 1 and 2).

Cellular composition of the TME

immune-checkpoint blockade *A type of anticancer immunotherapy that blocks checkpoint proteins on T cells or their targets on cancer cells and promotes an immune response and killing of the tumour.*

neoantigen *A tumour-specific antigen that is the result of a somatic coding mutation in the corresponding part of a gene. Antigens are protein fragments presented on the cell surface by the human leukocyte antigen (HLA) complex and recognised by different cells of the immune system.*

The TME has been studied extensively and, therefore, this section only presents a high-level summary of major cell types. Inflammation is a hallmark of cancer [Douglas Hanahan et al. 2011] and reflects the interplay of antitumorigenic and pro-tumorigenic immune responses (Figure 1.4). The antitumour response is primarily generated by cytotoxic $CD8^+$ T lymphocytes that recognise antigens presented on the tumour cell surface [Philip et al. 2022; Raskov et al. 2021]. The success of this response is dependent on priming and prior education by professional antigen-presenting cells (dendritic cells) and is facilitated by $CD4^+$ T helper cells [Borst et al. 2018; Waldman et al. 2020]. B cells have diverse functions, including antigen presentation, antibody production and cytokine-mediated cytolytic effects [Sharonov et al. 2020]. In many cancers, B cells form dense aggregates with T cells and dendritic cells to form tertiary lymphoid structures (TLSs) [Schumacher et al. 2022] that are generally associated with favourable cancer outcomes and responses to immune-checkpoint blockade [Sautès-Fridman et al. 2019]. Natural killer cells might have a particularly important role in anticancer immunity due to their ability to recognise stressed cells irrespective of neoantigen presentation [N. K. Wolf et al. 2023].

By contrast, CD4+ regulatory T (T_{reg}) cells can suppress the immune response

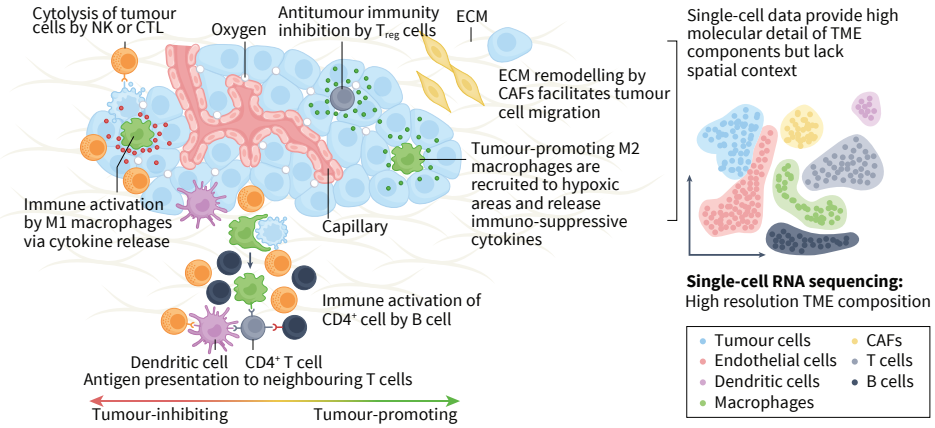


Figure 1.4: Cellular composition of the tumour microenvironment (TME) [Seferbekova et al. 2023]. The TME is composed of different cell types, involved in tumour-suppressing (on the left) and tumour-promoting (on the right) interactions with cancer cells. While TME composition can be captured by single-cell omic approaches (such as single-cell RNA sequencing), its spatial context is lost as a consequence of tissue dissociation. CAFs, cancer-associated fibroblasts; CTL, cytotoxic T lymphocyte; ECM, extracellular matrix; EMT, epithelial-to-mesenchymal transition; NK, natural killer; T_{reg}, regulatory T cell

through cytokine release (for example, IL-10) or by modulation of antigen presentation functions. In certain situations, some immune cells, such as macrophages, myeloid-derived suppressor cells and granulocytes, can actively promote tumour progression through two-way interactions with the tumour, other immune cells and the stroma [DeNardo et al. 2019; Goswami et al. 2023; Murdoch et al. 2008]. The stroma also plays an active role in shaping the functional immune response to cancer. A variety of studies have demonstrated that cancer-associated fibroblasts and myofibroblasts typically change shape, become more proliferative, undergo diverse transcriptional and metabolic changes, engage in angiogenesis, and interact with leukocytes and tumour cells while the resulting ECM changes in chemical composition and mechanical properties [Davidson et al. 2020; Kalluri 2016; Lendahl et al. 2022; Sahai et al. 2020; Valkenburg et al. 2018].

Cellular interactions and neighbourhoods

Although scRNA-seq studies have generated comprehensive catalogues of detailed TME cell types and states, these data only allow circumstantial insights into how cells interact. Spatial transcriptomic and proteomic technologies are

beginning to show that the cells constituting the cancer ecosystem self-organise into micro-anatomical neighbourhoods, which could provide insights into the frequency and nature of different cellular interactions (Figure 1.5) (Box 2). The majority of spatial omic cancer studies to date have used spatial transcriptomics as commercialised by 10x Visium (Box 1). These studies reveal spatially structured TMEs in breast [Andersson et al. 2021; Ståhl et al. 2016; S. Z. Wu et al. 2021], prostate [Berglund et al. 2018], pancreatic [Moncada et al. 2020], colorectal [Qi et al. 2022], skin [Ji et al. 2020], brain [Ravi et al. 2022], liver [R. Wu et al. 2021], bladder [Gouin et al. 2021] and other cancer types [Barkley et al. 2022; Erickson et al. 2022]. The ability to use whole-transcriptome readouts also helped establish the presence and broad co-localisation of rare cell types (Figure 1.5). However, the current super-cellular resolution (55 µm, 3–10 cells) requires informatic deconvolution of the measured signals in each spot.

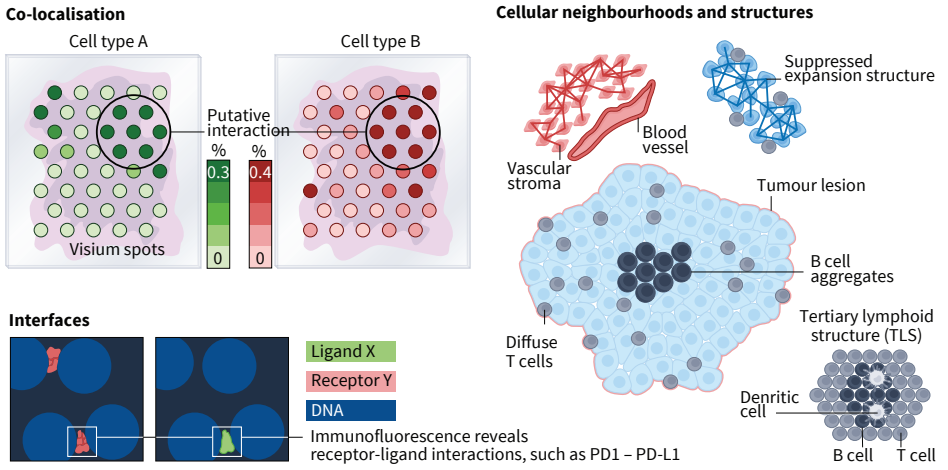


Figure 1.5: Cellular interactions and neighbourhoods [Seferbekova et al. 2023]. Cellular interactions: spatial co-localisation and concomitant expression of receptors and ligands are indications of cellular interactions. Cellular neighbourhoods: spatially resolved proteomic studies detect clinically relevant neighbourhoods of molecularly related cell types [Danenberg et al. 2022]. Cellular interfaces: subcellular resolution imaging reveals receptor–ligand localisation at cellular interfaces [Nirmal et al. 2022]

Highly multiplexed antibody detection approaches have been used to generate single-cell resolution readouts of the spatial distribution of the major cell types in breast cancer, colon cancer and melanoma, amongst others [Danenberg et al. 2022; Jackson et al. 2020; Nirmal et al. 2022; Schürch et al. 2020]. These approaches permit the detection of microscopic neighbourhoods with distinct composition of

tumour, immune, stromal and vascular cells. In two studies that analysed hundreds of breast cancers by applying imaging mass cytometry (Box 1) to tissue microarrays, several recurrent neighbourhoods were identified [Danenberg et al. 2022; Jackson et al. 2020] (Figure 1.5). The biological feasibility of the identified neighbourhoods was provided by their relationship with well-defined clinical subtypes and survival outcomes. For example, consistent with a failing antitumorigenic response, 'suppressed expansion structures' contained many dysfunctional T and T_{reg} cells and were enriched in breast cancers with poor clinical outcomes. By contrast, a 'vascular stroma' signature was enriched in a very favourable outcome subgroup of hormone receptor-positive breast cancers (Figure 1.5) . The spatial patterns formed by individual cell types were also examined. In breast cancers, whereas T cells were diffusely scattered, B cells tended to form aggregates, were in multiple neighbourhood types, and were associated with disparate clinical subtypes and clinical outcomes, suggesting that B cell function might be dependent on the precise TME context [Wieland et al. 2021] (Figure 1.5). However, some tumour phenotypes, such as cell proliferation, can be shaped by further long-range interactions [Gaglia et al. 2022].

The proximity of cells does not necessarily equate to an interaction. To address this limitation, a combination of CyCIF (Box 1) and high-resolution three-dimensional optical deconvolution was used to study the physical features that might be associated with cellular interactions [Nirmal et al. 2022]. In melanoma tissues, probable interactions were identified amongst approximately 20% of immune cells at the tumour–stroma interface (Figure 1.5). These interactions were often complex, involving three or more cells (for example, a melanoma cell contacting two CD8+ cytotoxic T lymphocytes and one $CD4+$ T_{reg} cell) with associated cell surface molecular polarisation and receptor–ligand interaction. In some cases, cellular processes from macrophages extended more than 10 μm to make contact with other cells. Standard fine sections would miss many of these interactions, whereas basic proximity measurements would tend to overestimate the nature of interactions. As discussed in Box 2, although these analyses remain technically challenging and have low throughput, they demonstrate the unique potential afforded by spatial omics for the mapping of functional cellular interactions in tissue space.

Box 1: Enabling technologies

A range of spatially resolved approaches for mapping DNA, RNA and proteins has been developed [Lewis et al. 2021; Rao et al. 2021]. Higher spatial resolution typically implies smaller molecular quantities. Other important parameters are the field of view, which ranges from millimetres squared to centimetres squared, and method sensitivity.

Macrodissection and microdissection

The foundations were laid by macrodissection followed by multi-omic profiling but new technologies offer a microscopic and single-cell resolution. In multi-region sequencing [Gerlinger et al. 2012; Yates et al. 2015], the collection of several biopsy samples from multiple geographically separated tissue regions is coupled with later bulk sequencing of DNA or RNA. laser capture microdissection (LCM) allows the profiling of even smaller regions of interest using laser ablation [Emmert-Buck et al. 1996]. LCM is very versatile as it can be combined with high throughput sequencing, including whole genome sequencing (WGS), and single-cell RNA or DNA sequencing [Casasent et al. 2018; Cui Zhou et al. 2022; Ellis et al. 2021; Nichterwitz et al. 2016].

Spatial molecular barcodes

This set of methods utilizes probes with unique barcode sequences to capture molecules and then map them to specific spatial locations. So far, these technologies have mainly focused on transcriptomic profiling with spatial transcriptomics [Ståhl et al. 2016], 10x Visium and Slide-seq [Rodriques et al. 2019; Stickels et al. 2021] being the most popular protocols among many others [Cho et al. 2021; Liu et al. 2020; Vickovic et al. 2019]. Spatially resolved DNA sequence capture [Zhao et al. 2022] has enabled spatially resolved genome-wide copy number calling along with several new methods for spatial epigenomic profiling [Deng, Bartosovic, Kukanja et al. 2022; Deng, Bartosovic, S. Ma et al. 2022]. Although these methods support unbiased discovery and result in data with high molecular content, the resolution is limited by the diameter of each spot and currently does not allow profiling at the single-cell level or point mutation detection. Further improvements in resolution are currently being developed [A. Chen et al. 2022; Cho et al. 2021; Vickovic et al. 2019].

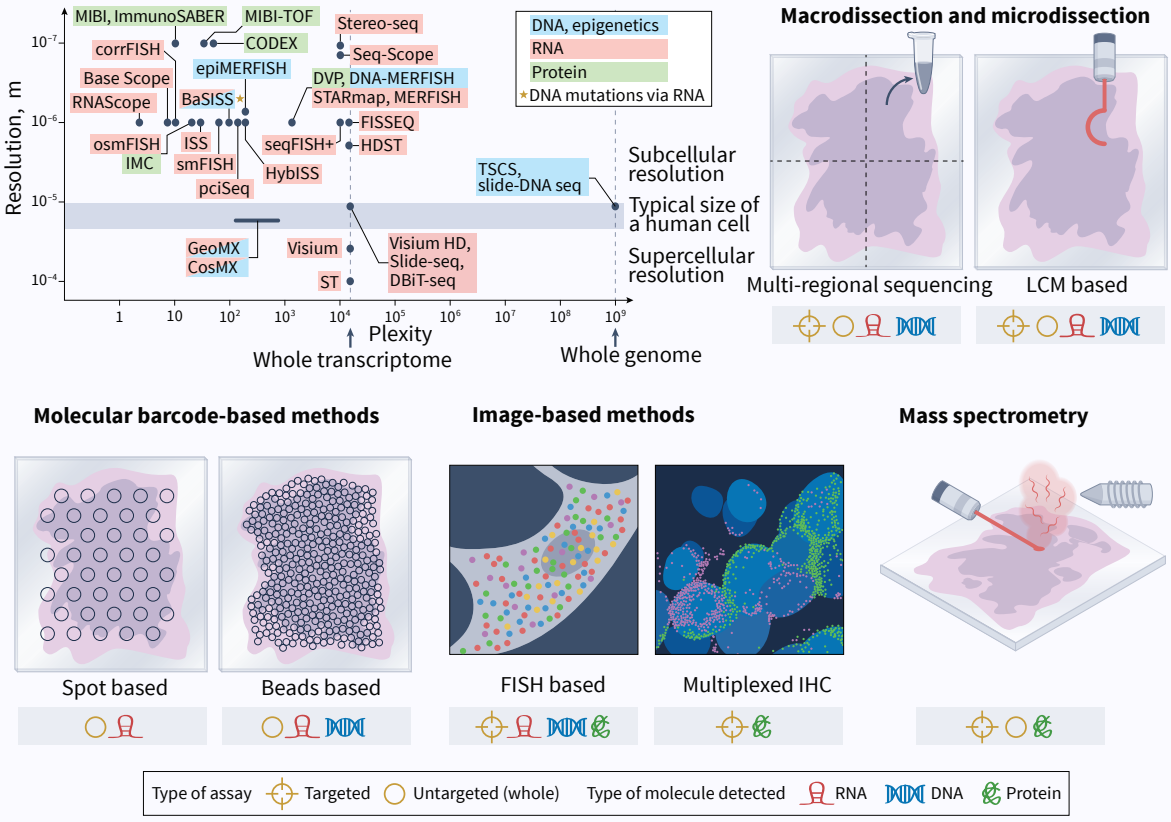
Image-based technologies

This group of methods involves *in situ* hybridisation of fluorescently labelled probes (FISH) to nucleic acids of interest to identify their location with single-cell or subcellular resolution. A variety of multiplexed extensions exists [Alon et al. 2021; F. Chen et al. 2016; K. H. Chen et al. 2015; X. Chen et al. 2018; Codeluppi et al. 2018; Coskun et al. 2016; Gyllborg et al. 2020; J. H. Lee et al. 2015; Lyubimova et al. 2013; Qian et al. 2020; X. Wang et al. 2018], including multiplexed error-robust FISH (MERFISH) [T. Lu et al. 2022; J.-H. Su et al. 2020], sequential FISH (seqFISH) [Eng et al. 2019], RNAscope [F. Wang et al. 2012] and *in situ* sequencing (ISS) [Ke et al. 2013] for spatial transcriptomic profiling. The latter two protocols have been further adapted to permit mutation detection using the Base Scope [Baker et al. 2017] and BaSISS protocols [Lomakin et al. 2022]. immunohistochemistry (IHC) or immunofluorescence uses antibodies to localise proteins or other molecules. Highly multiplexed spatial proteomics with dozens of targets can be achieved using DNA barcodes [Goltsev et al. 2018; Saka et al. 2019], multiplex immunofluorescence (CyCIF) [Jia-Ren Lin et al. 2018] and other approaches [Gerdes et al. 2013; Keren et al. 2019; Jia-Ren Lin et al. 2015]. Multi-omic approaches, such as GeoMx or CosMx [Bergholtz et al. 2021; He 2022], allow researchers to assess multiple layers of biological information within one cell simultaneously. Image-based technologies provide high-resolution tissue profiling but require a pre-defined panel of targets.

Mass spectrometry

A further class of methods combines metal-conjugated antibodies and mass spectrometry.

Spatial resolution is provided by spatial laser ablation (imaging mass cytometry (IMC)) [Giesen et al. 2014] or ion beams (multiplexed ion beam imaging (MIBI)) [Angelo et al. 2014]. These approaches detect proteins and smaller molecules, such as lipids, metabolites, and drugs, in single cells [Giesen et al. 2014], as reviewed in Mund, Brunner et al. [2022]. Deep Visual Proteomics (DVP) combines artificial intelligence-driven LCM with mass spectrometry to map protein abundances onto tissue slides [Mund, Coscia et al. 2022].



Tumour histology and macrostructure

The histological appearance of tumour cells and their micro-anatomical structures also reflect the underlying genomic alterations as revealed by two recent pan-cancer analyses using artificial intelligence-based digital histopathology [Fu et al. 2020; Kather et al. 2020]. In some cases, histological patterns have been found to be indicative even of distinct subclonal alterations [Coudray et al. 2018; Loeffler et al. 2022; Lomakin et al. 2022]. It is well established that tumours exhibit characteristic histological structures indicative of their natural history (Figure 1.6). Many cancers exhibit adjacent areas of carcinoma *in situ*, which are considered to be precursor lesions. This phenomenon is well established in melanoma [Nirmal et al. 2022] and many adenocarcinomas: in breast cancer up to two-thirds of cases have histologically distinct areas of ductal carcinoma *in situ* [Kole et al. 2019], around 15% of colorectal adenocarcinomas exhibit adjacent adenomas [Ponz de Leon et al. 1990], and 45% of oesophageal adenocarcinomas show detectable areas of its precursor lesion Barrett oesophagus [Sawas et al. 2018]. Interestingly, the phylogenetic and histological relationships between precursor lesions and invasive cancer are not always aligned with patterns indicative of branching [Ross-Innes et al. 2015; Stachler et al. 2015] or multiclonal invasion [Casasent et al. 2018]. Therefore, spatial genomic and transcriptomic analyses using histologically informed dissection allow for the exploration of mechanisms and phenotypes of disease progression in greater detail, revealing subclone and histology-specific changes [Lomakin et al. 2022].

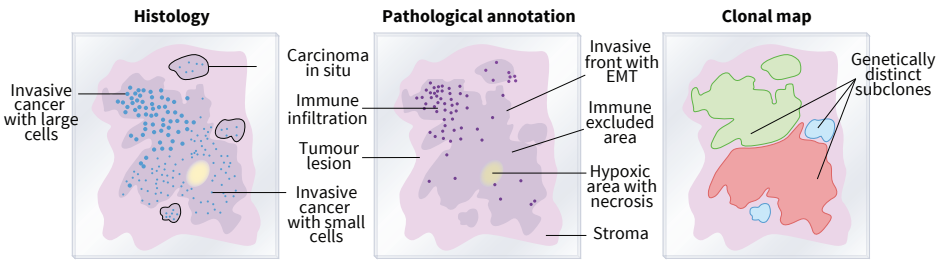


Figure 1.6: Tumour macrostructure [Seferbekova et al. 2023]. Different tumour regions can be identified in some tumours based on histological staining. Such histopathological findings can be supplemented and underpinned by detailed molecular maps

At the scale of millimetres, cancers are often described in terms of a tumour core that is separated from the surrounding stroma by an invasive margin (Figure 1.6). Transcriptional diversity has been reported across these landscapes [Ber-

glund et al. 2018], with hypoxia and EMT-like signatures localising to the core and invasive margin, respectively [Puram et al. 2017; Thomlinson et al. 1955]. The core–stroma distinction is fundamental to common classifications of immune infiltrates that correlate with cancer subtype, clinical outcome and therapy response [Galon et al. 2014; Hammerl et al. 2021; Keren et al. 2018]. At one end of the immune infiltrate continuum are inflamed tumours, where plentiful immune cells infiltrate the tumour core, at the other end are immune deserts that are devoid of infiltrates, and between these are immune-excluded tumours where immune cells are largely restricted to the neighbouring stroma [Hegde et al. 2020]. However, recent spatial studies in pancreatic and breast cancer reveal that these immune characteristics can be highly localised and coexist in different parts of the same tumour [Danenberg et al. 2022; Grünwald et al. 2021; Tavernari et al. 2021]. A challenge will therefore be to decipher how localised variation in immune response shapes the evolutionary outcome of the tumour.

1.4. Evolution of the cancer ecosystem

The cancer ecosystem detailed in the previous section changes the dynamics of cancer evolution as tumour cells face selection pressures that maximize growth-promoting and minimize growth-inhibiting interactions. Foremost are cell–cell and other local interactions with the immune system but also with fibroblasts, vasculature and even neurons, which can lead to the selection of specific genomic alterations (Figure 1.7). The fact that the TME is itself spatially heterogeneous suggests that tumour cells experience localised variations in selective pressure.

Evolution of immune evasion

Various parts of the immune system constrain tumour growth. The observed range of associations between specific genetic alterations and the immune system cells of the TME [Rooney et al. 2015; Thorsson et al. 2019] illustrates the presence of selective pressures on cancer evolution. Neoantigens, which derive from coding mutations in cancer genomes, have been reported to elicit T cell reactivity in lung cancer [McGranahan et al. 2016]. In tumours with greater cytolytic activity, loss of neoantigen presentation capabilities was more frequent, which is indicative of greater selective pressures [Rooney et al. 2015; Shukla et al. 2015]. Other signs of selection for immune evasion include mutations in *CASP8*, which serve as a potential way to resist apoptosis induced by FASL [Rooney et al. 2015]. Furthermore, gains of immunosuppressors, such as CD274 (also known as PDL1),

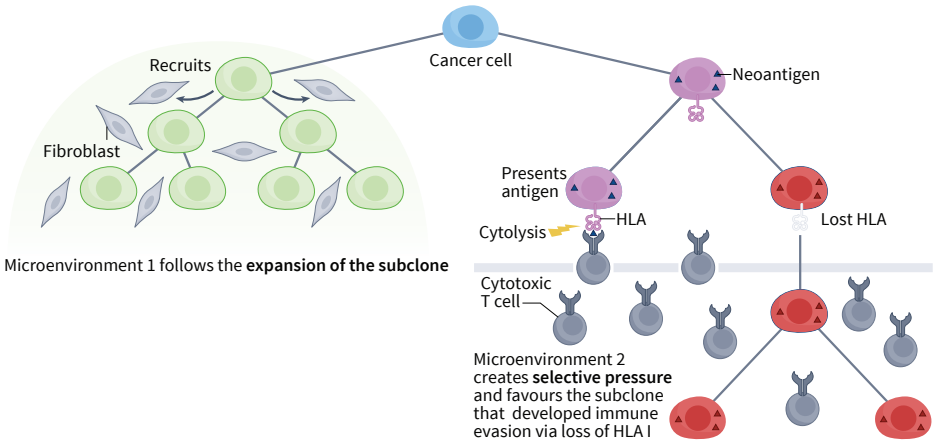


Figure 1.7: Co-evolution of tumour cells and microenvironments [Seferbekova et al. 2023]. Some tumour subclones attract characteristic tumour microenvironments (TMEs), which subsequently spatially overlap with the clonal territories. Alternatively, microenvironments may select for specific tumour characteristics such as immune evasion; here, immune evasion is shown as impaired antigen presentation due to loss of expression of human leukocyte antigen 1 (HLA1).

are widespread among Hodgkin lymphomas and are detected in 0.7% of solid tumours [Goodman et al. 2018].

immune evasion *The process by which tumour cells develop several mechanisms that help them to continue to grow and expand by escaping immune control. Immune evasion is thought to be typically preceded by two other phases: elimination, when the immune system recognises and eliminates tumour cells, and equilibrium, when the pressure from the immune system stalls tumour growth and expansion, and negative selection takes place.*

The signs of immune evasion increase during later stages in cancer evolution. No evidence of negative selection was reported in diploid genomes of normal tissues and among clonal alterations in primary cancers, indicating that there is no depletion of coding mutations during normal evolution [Martincorena et al. 2017; Van den Eynden et al. 2019]. Yet, the signs of selection for immune evasion become stronger after clonal expansion at the subclonal level. For example, immune evasion via deletions of human leukocyte antigen (HLA) class I alleles, predominantly subclonally and in metastases, has been described in lung cancer [McGranahan et al. 2017] and multiple other cancer types [Watkins et al. 2020] (Figure 1.7). This mode of evolution is plausible as it may require a certain tumour size for the immune system to recognise its neoantigens; at this stage, however, a mutation enabling escape from these new selective pressures arises in a single cell of the tumour and seeds the growth of a tumour subclone. While it is possible for it to replace all other tumour cells, the timing at which this occurs depends on the strength of immune control and its escape rate.

negative selection *The removal of deleterious variants from the population.*

Furthermore, selective pressures of the immune system determine prognosis and immune evasion is found in relapses. It has been reported that strongly

immunogenic neoantigens are prognostically favourable in pancreatic cancer [Balachandran et al. 2017] but are preferentially lost at recurrence [Łuksza et al. 2022]. In melanoma, mutations of *JAK1* or *JAK2*, which inhibit cytolysis, and in *B2M*, which reduce antigen presentation, are often found in samples that acquired resistance to immune-checkpoint inhibitors [Sade-Feldman et al. 2017; Zaretsky et al. 2016].

Maps of cancer–TME interactions

Establishing associations between genomic alterations and the microenvironment can be challenging due to the large genomic heterogeneity between primary cancers. Conceptually, spatial genomics offers new ways of analysing such interactions at high spatial resolution and also within the same tumour. For intratumour diversity, the observations occur in an isogenic background and the set of genomic differences between subclones is small compared to the large intertumour diversity (Figure 1.8). A proof of concept has been established in breast cancer, where various degrees of immune infiltration were found at the level of cancer subclones [Lomakin et al. 2022]. Interestingly, such observations could also be linked to the stage of invasion, with some TME regions defined by the subclonal lineage and others mostly by histological progression from carcinoma *in situ* to invasive cancer. Direct experimental evidence was provided by Perturb-map, a mouse *in vivo* spatial CRISPR screen, which revealed that specific gene knockouts lead to characteristic changes in the immune microenvironment. For example, it was shown that T cell infiltration is increased by *Socs1*$^{-/-}$ knockout in the tumour but decreased by *Tgfbr2*$^{-/-}$ knockout [Dhainaut et al. 2022] (Figure 1.8).

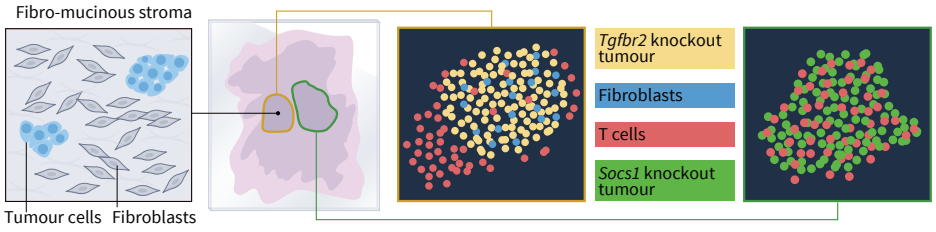


Figure 1.8: Co-evolution of tumour cells and microenvironments [Seferbekova et al. 2023]. Examples of experimentally observed associations between clones and TMEs. The yellow *Tgfbr2*$^{-/-}$ knockout tumour clone exhibits a fibroblast-enriched stroma with no presence of T cells, whereas the green *Socs1*$^{-/-}$ knockout clone has a stroma infiltrated with T cells [Dhainaut et al. 2022].

Further cell types beyond immune cells have been associated with cancer evolution. Fibroblasts are an extremely versatile group of cells that perform various roles that are exploited during tumour growth. For example, cancer-associated fibroblasts can be educated by cancer cells with gain-of-function mutations in *TP53* to create a pro-tumorigenic microenvironment [Vennin et al. 2019]. Furthermore, fibroblasts can amplify tumour-intrinsic oncogenic signalling in $KRAS^{G12D}$-mutant pancreatic adenocarcinoma [Tape et al. 2016]. Spatial genomics adds to these observations. In breast cancer, imaging mass cytometry (Box 1) revealed enrichment of different fibroblast and myofibroblast populations in *TP53*-mutant and genomically unstable cancers [Ali et al. 2020]; however, the nature of those associations is unclear. The aforementioned Perturb-map screening revealed that $Tgfbr2^{-/-}$ knockout created a fibro-mucinous stroma characterized by TGFβ-activated fibroblasts and T cell exclusion [Dhainaut et al. 2022].

Another cell-extrinsic and environmental factor influencing cancer evolution is hypoxia, the often localised depletion of oxygen in tumours; it is well established as a marker of poor outcomes [Harris 2002] and has been shown to select for certain mutations such as *TP53* [Graeber et al. 1996]. A recent study combining scRNA-seq and immunofluorescence found that hypoxic microniches contain quiescent cancer cells resistant to T cell attack [Baldominos et al. 2022].

Lastly, spatial genomics may also shed light on how different cancer subclones interact with each other. Reports of clonal cooperation, observed in heterologous xenotransplants, date back to at least 1989 [Miller et al. 1989]. Elements of clonal cooperation have since been described as involving direct and microenvironmentally mediated interactions [Alonso-Curbelo et al. 2021; Cleary et al. 2014; Tammela et al. 2017; J. B. Williams et al. 2020; Hengbo Zhou et al. 2017]. The ability to comprehensively map genomic evolution, subclone-specific gene expression and TMEs is thus likely to further elucidate the full extent and mechanisms of this phenomenon.

Disseminated tumour cells, metastatic niches and organotropism

Metastases are seeded by disseminated tumour cells (DTCs). The mechanisms that determine whether any given DTC survives, enters a dormant state or forms a metastasis are poorly defined, but it is evident that these cells can exploit a variety of existing cellular interactions and spatial microenvironments. For example, dissemination was shown to depend on environmental triggers tied to the circadian rhythm [Diamantopoulou et al. 2022]. In breast cancer, DTCs physically associated with neutrophils were shown to exhibit a more aggressive phenotype

[Szczerba et al. 2019]. Additionally, aggregates or even fusions of DTCs and macrophages have been described, further illustrating that metastatic dissemination may also rely on a range of cellular interactions [Adams et al. 2014].

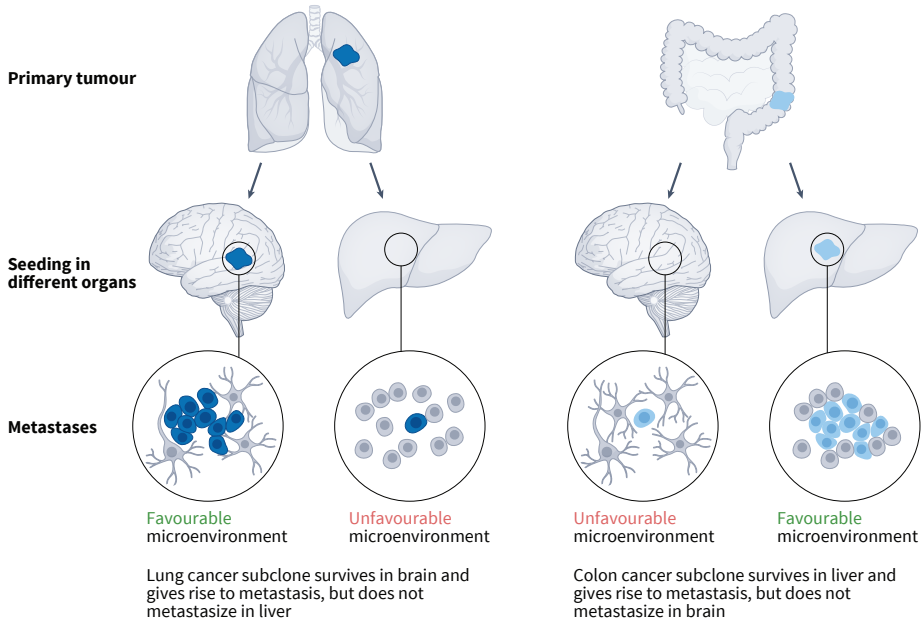


Figure 1.9: Organotropism and the seed and soil hypothesis of metastasis [Seferbekova et al. 2023]. Cancer cells from the primary tumour migrate throughout the body via blood vessels, with some of them able to seed metastatic lesions in a distant organ. The target TME determines the fitness of clones at the site of metastasis favouring some tumour lineages over others.

The existence of organotropism, whereby certain cancer types and specific molecular alterations exhibit distinct patterns of metastatic organ involvement, is well established and is often considered in terms of the classical seed and soil hypothesis suggested by Paget, which posits that certain tissues provide host environments susceptible to specific tumour types [Fidler 2003; Paget 1889] (Figure 1.9). Organotropism seems to be mostly determined by the cell of origin of the primary tumour, the ability of DTCs to interact with the metastatic host environment and the anatomical proximity of certain target organs such as liver metastases in colorectal cancer. In breast cancer, brain metastases were shown to be facilitated by the interactions with glutamatergic neurons [Zeng et al. 2019], and a similar pattern was also observed in primary brain cancers where neoplastic cells form networks with normal astrocytes via synapses [Venkataramani et al. 2022].

The notion that both intrinsic and microenvironmental features of cellular populations in the primary tumour influence metastatic rates is further supported by lineage-tracing studies in mouse models [Quinn et al. 2021]. However, a genomic element also emerges. A recent pan-cancer analysis of 25,000 metastases of 10 cancer types revealed 57 genetic associations of organotropism within the same primary tumour type [B. Nguyen et al. 2022]. Although the presence of many genomic alterations was associated with increased rates of metastasis in multiple organs, there were also examples of genomic alterations associated with reduced burden, for example, *RBM10* mutations being less prevalent in brain metastases of lung cancers. However, it is worth noting that these genetic patterns appeared to be mostly specific to certain primary tumour sites rather than to target sites.

A series of studies reported that tumours prime target tissues to create pre-metastatic niches facilitating colonisation and metastatic outgrowth [Peinado et al. 2017]. This occurs prior to or on the arrival of DTCs through soluble factors and extracellular vesicles released by the primary tumour that can alter the systemic immune response and induce localised regions of vascular and stromal remodelling in target organs. Furthermore, DTCs can enter dormant niches that facilitate long-term survival and even evasion of therapy. In prostate cancer, DTCs can compete with haematopoietic stem cells for the occupation of established endosteal niches [Shiozawa et al. 2011], whereas perivascular niches have been observed in bone, lung and brain [Ghajar et al. 2013; Kienast et al. 2010]. In melanoma, age-dependent changes of fibroblasts have been reported to induce emergence from dormancy in lung deposits [Fane et al. 2022]. Similarly, in a mouse model of melanoma, metastases in lymph nodes were found to promote immune evasion and facilitate further metastasis in distant organs [Reticker-Flynn et al. 2022]. As all of these phenomena involve cellular and spatial interactions, spatial genomic, transcriptomic and proteomic analyses are likely to provide further insights into these processes.

1.5. Translational opportunities

As discussed in the preceding sections, spatial genomic methods generate rich insights into the cellular composition, organisation and interactions that shape the evolving cancer ecosystem. For any given cancer, the genome bears the scars of its unique, past evolutionary journey; however, clinical interventions are designed to shape the future disease course (Figure 1.10). A major hope for cancer care is therefore that evolutionary features might be leveraged to better predict

the clinical trajectory. Because tumour evolution is a spatial process, spatially resolved features may provide additional means to predict outcomes. Furthermore, a better understanding of the cellular interactions within the ecosystem might identify potential therapeutic vulnerabilities that could be exploited to control the evolutionary process [Andersson et al. 2021; Ji et al. 2020; Maldegem et al. 2021; Moldoveanu et al. 2022; Moncada et al. 2020]. Initial studies hint towards the opportunities afforded by these approaches for novel biomarker discovery [Danenberg et al. 2022; Keren et al. 2018; Moldoveanu et al. 2022], in understanding the clinical relevance of genetic subclonal structure [Erickson et al. 2022; Lomakin et al. 2022; Zhao et al. 2022] and in identifying therapeutic vulnerabilities through a deepened understanding of cellular interactions in both human samples [Ji et al. 2020; Keren et al. 2018; Moncada et al. 2020] and preclinical animal models [Dhainaut et al. 2022].

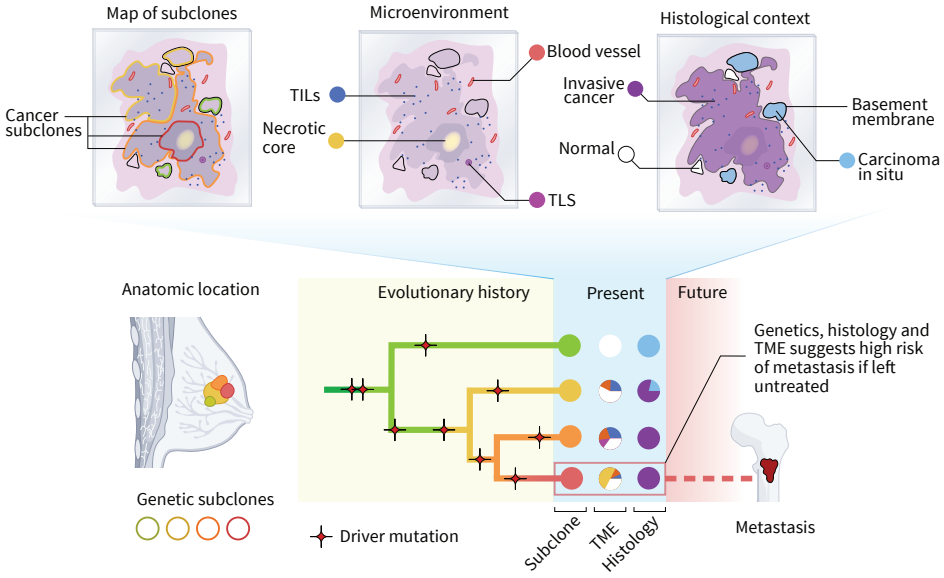


Figure 1.10: Translational opportunities [Seferbekova et al. 2023]. Subclone maps, tumour microenvironment (TME) composition and histological context each provide clinically and prognostically relevant features. Integrating these levels of information may help identify ill-fated clonal lineages that may not be found using either component alone. For example, the metastatic potential of two invasive subclonal lineages may differ based not only on the portfolio of their driver mutations but also on the TMEs that they inhabit. tumour infiltrating lymphocyte (TIL)s, tumour infiltrating lymphocytes; TLS, tertiary lymphoid structure.

Box 2: Spatial omic analysis

A typical tumour cross-section covers several centimetres of tissue containing up to 1,000,000 cells. Spatial omic methods characterize the molecular features of these cells with a resolution as fine as 0.2 µm, depending on the method. The wealth and detail of data bring a series of data analysis challenges [Palla et al. 2022] (see the figure).

Raw signal analysis

In the case of molecular barcode-based technologies (Box 1), primary steps of raw data processing include aligning and counting reads or molecular barcodes per genomic region of interest, with sequencing barcodes defining the spatial sector. Image-based methods, by contrast, utilize fluorescent signals, often measured across multiple fluorescence channels and cycles. These signals must be detected, quantified and decoded to their corresponding target molecule, which can be defined by a specific combination of colours across cycles. For whole-slide imaging, fields of view need to be stitched together into one tissue image [J - R Lin 2021]. Background fluorescence, blur and signal crowding can negatively affect the ability to decode signals [Dries et al. 2021; Gataric et al. 2021; Hickey et al. 2022; Ke et al. 2013; Perkel 2019]. Technologies with a super-cellular resolution require cell decomposition algorithms, where known cell types usually obtained from the single-cell data serve as a reference [Kleshchevnikov et al. 2022; Lomakin et al. 2022; Moncada et al. 2020; Nieto et al. 2021]. Signals with a subcellular resolution are spatially aggregated either by clustering [Park et al. 2021; Petukhov et al. 2022; Prabhakaran 2022; Soldatov et al. 2019] or based on cell segmentation derived from nuclear DAPI signals, membrane markers or haematoxylin and eosin (H&E) staining [Lohoff et al. 2022; Petukhov et al. 2022; Qian et al. 2020; Schapiro et al. 2022].

Data integration

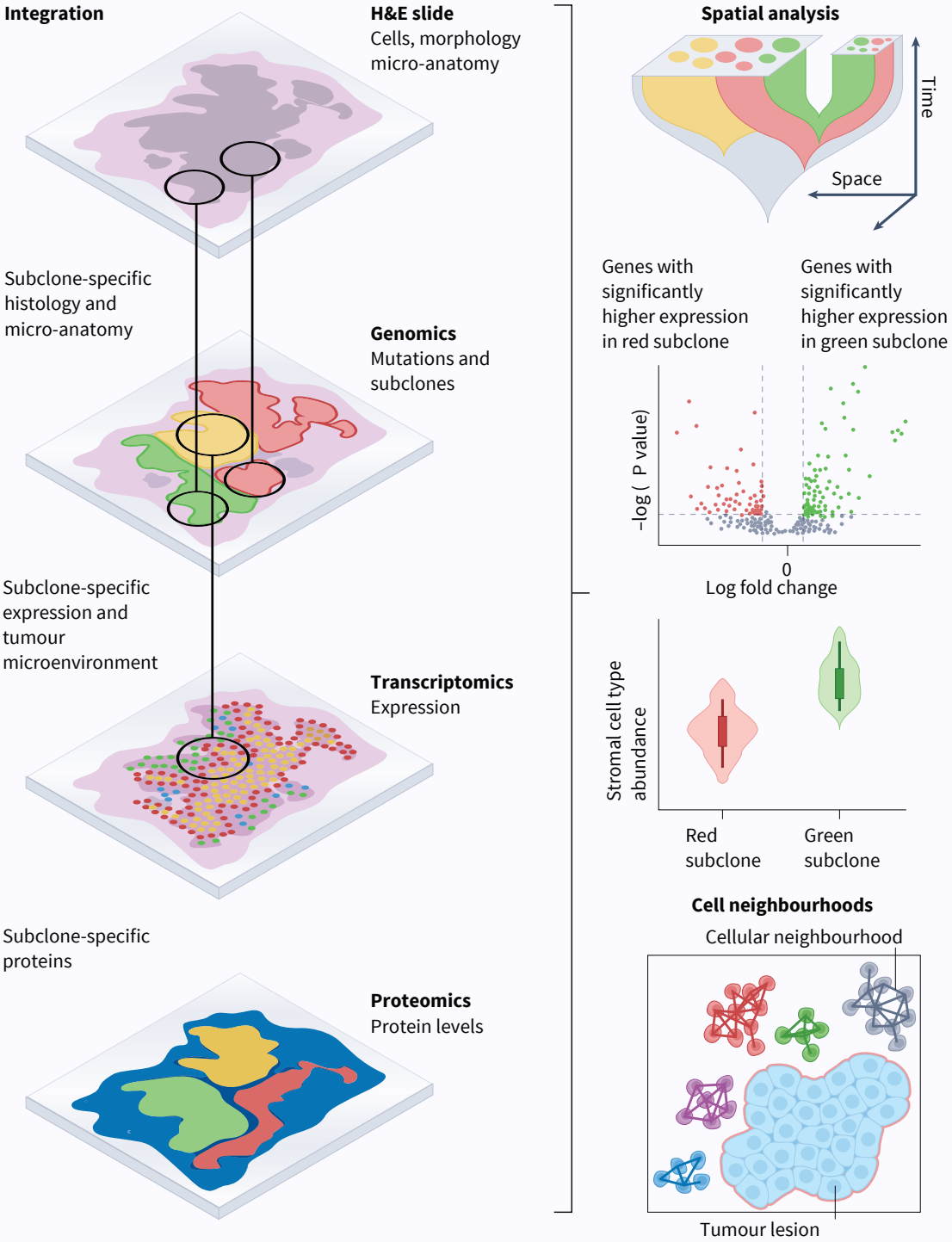
Single-cell resolved data can be mapped [Zhang 2022] to deconvolve larger spots or to annotate known cell types. Single-cell whole transcriptomes or multi-omics can be further mapped into space, leveraging a smaller common set of reference genes that are interrogated spatially [Stuart et al. 2019; Zhang 2022]. Another option is to integrate multiple spatially resolved modalities from consecutive tissue sections. Alignment of distorted sections and 3D reconstruction of a tissue block can be done using image registration algorithms [Kiemen et al. 2020]. Multimodal characterisations of microscopically matched tissue areas, including spatial transcriptomics, proteomics and histopathology, are also possible [Lomakin et al. 2022; Velten et al. 2022].

Spatial statistics

Spatial differential expression models enable the description of novel cell states showing distinct spatial patterns that are difficult to recover from single-cell data [Lopez et al. 2022; Svensson et al. 2018]. A group of cells may appear enriched in a recurrent way forming distinct neighbourhoods [Danenberg et al. 2022; Dong et al. 2022; McCaffrey et al. 2022; Nirmal et al. 2022], a pattern which could be picked up by various topic and graph-based models. Mechanistic models attempt to explain expression plasticity mediated by cell–cell interactions and signalling at the cell interface [Arnol et al. 2019].

Spatial lineage tracing

Spatial RNA or DNA sequencing technologies provide wide but low genomic coverage, which limits lineage tracing to copy number alterations and a few point mutations [Elyanow et al. 2021; Nagasawa et al. 2021; Zhao et al. 2022]. Current protocols also have a relatively low spatial resolution (Box 1). Microscopy methods target specific point mutations and require prior knowledge of the clonal structure. Yet, they achieve a higher spatial resolution to map a more fine-grained clonal structure across wide fields of view [Lomakin et al. 2022]. The incremental nature of branching evolution also provides opportunities to associate the genetic changes on each branch with phenotypic changes.


Integration
H&E slide
Cells, morphology
micro-anatomy
Subclone-specific
histology and
micro-anatomy
Genomics
Mutations and
subclones
Subclone-specific
expression and
tumour
microenvironment
Transcriptomics
Expression
Subclone-specific
proteins
Proteomics
Protein levels
Spatial analysis
Time
Space
Genes with
significantly
higher expression
in red subclone
Genes with
significantly
higher expression
in green subclone
–log (P value)
0
Log fold change
Stromal cell type
abundance
Red
subclone
Green
subclone
Cell neighbourhoods
Cellular neighbourhood
Tumour lesion

Spatial biomarker discovery

Spatial biomarkers predict clinical outcomes by leveraging information about cellular organisation or intercellular relationships. In its most simplistic form, this approach might measure the distribution of a certain cell type in relation to another. This is the basic concept of commonly used, histomorphological scoring of TLSs or tumour infiltrating lymphocytes (TILs) within the tumour or stromal compartment (Figure 1.5). Stromal TILs are associated with a superior prognosis and response to chemotherapy or immune-checkpoint inhibitors in triple-negative breast cancer, non-small-cell lung cancer and melanomas [Azimi et al. 2012; B. Chen et al. 2020; Helmink et al. 2020; H. J. Lee et al. 2016; Salgado et al. 2015]. However, in isolation, the reliability of TILs as a biomarker is hampered by heterogeneity and irreproducibility [Salgado et al. 2015]. Recent studies that use highly multiplexed, *in situ*, proteomic-based assays have started to dissect this complexity by characterizing TIL subtypes, expression, cellular colocalisation and neighbourhood patterns, identifying prognostically meaningful TIL-related patterns [Keren et al. 2018; Moldoveanu et al. 2022]. In general, head-to-head clinical trials are needed to confirm whether multiplex biomarkers carry additional prognostic and predictive value compared to single biomarkers such as PDL1 immunohistochemistry [S. Lu et al. 2019]. Notably, the Immunoscore [Galon et al. 2006] is a digital pathology-based assay that measures CD3+ and CD8+ lymphocyte density in the tumour core and invasive margin and has been proven to surpass clinical staging systems for disease-free survival prediction, demonstrating that higher-dimension datasets could be clinically transformative [Pagès et al. 2018].

Beyond a role in refining histology-based biomarkers, spatial omic approaches can also facilitate de novo biomarker discovery. This was demonstrated by two breast cancer studies (discussed in the Section 1.3) that extracted recurrent cellular communities from multiplex proteomic and morphological data using machine learning algorithms [Danenberg et al. 2022; Jackson et al. 2020]. Another use of multiplexed tissue imaging was an ovarian cancer study, which found that interactions of exhausted CD8+ T cells, PDL1+ macrophages and PDL1+ tumour cells were linked to treatment response [Färkkilä et al. 2020]. In these studies, the power to detect clinically meaningful features was derived from the fact that hundreds of samples could be analysed at high throughput by using tissue microarrays. However, the trade-off was that only a tiny part of the main tumour mass was sampled (core needle biopsies with a diameter of 0.6–0.8 mm), whereas the invasive margin and stroma were largely unsampled. An important next step for

spatial molecular analyses is to derive catalogues of typical cancer ecosystems at a tissue-wide scale and at single-cell resolution. These reference sets could serve as a basis for designing more focused studies that might take advantage of lower-plex, higher-throughput or rational subsampling approaches. The scalability and reproducibility of omic data will be critical factors for successful implementation within large-scale clinical studies.

Clinically relevant subclones

An unanswered question for personalized medicine is how to deal with subclonal driver mutations that can be targeted by antitumour therapy. By integrating additional layers of spatial data, it is possible to establish the spatial context and thus the clinical relevance of subclonal alterations [González-Silva et al. 2020; Lomakin et al. 2022] (Figure 1.10). For example, systemic therapies administered after primary breast cancer surgery are directed towards covert metastatic disease, with the intention of eliminating these cells before metastatic deposits have a chance to form. As we cannot directly assay disseminated tumour cells, treatment paradigms are based upon the properties of the primary tumour. It is reasonable to assume that a subclonal mutation within the invasive cancer might also be carried by disseminated tumour cells and could drive metastasis formation. By contrast, if the mutation is entirely restricted to precursor lesions, and hence a clone that has not yet even exhibited invasive capacity, we can be relatively confident that it will not be a driver of metastasis.

Integrating spatial transcriptomic or proteomic data provide further insights into the TME and the gene expression properties of subclones. This additional characterisation will be helpful for deriving predictions of the functional relevance of a particular subclonal driver mutation, such as by considering a range of intrinsic and extrinsic phenotypes, including vascular invasion or a high-risk TME niche (Figure 1.10). Although further research is needed, it is possible that spatial analyses could be used in early detection settings or to complement histopathology-determined completeness of excision, with the latter application yielding opportunities for biomarker-driven surgical and adjuvant radiotherapy clinical trials.

Preclinical studies

Although spatial genomic analysis enables reconstruction of the evolutionary past and prediction of the clinical future, spatial genomic experiments in an-

imal models can add direct mechanistic evidence of the roles of particular cancer-associated mutations and/or the effects of spatial aspects of the cancer TME. Such experiments are expected to provide clinically important insights into the extracellular consequences of perturbations in cell-intrinsic processes, with many opportunities for biomarker discovery [Janiszewska et al. 2019; Maldegem et al. 2021]. A particularly exciting approach is the combination of spatial profiling with *in vivo* CRISPR screens [Dhainaut et al. 2022; Ji et al. 2020]. Using this approach, both cell-intrinsic phenotypic and local TME associations of tens of genetically distinct clones can be studied simultaneously in a single animal, leading to insights into the precise mechanisms that lead to diverse clinical fates of genetically divergent clones that nonetheless share many genetic and host characteristics [Dhainaut et al. 2022]. These models could lead the way in developing pan-ecosystem therapeutic strategies that simultaneously target cell-intrinsic vulnerabilities and harness the properties of the TME in an attempt to terminate cancer evolution.

1.6. Conclusions and future perspectives

While it is widely accepted that the population genetics of carcinogenesis can be well described in terms of somatic evolution, the spatiotemporal details of the process are not fully understood. Open questions remain over the extent to which cellular interactions with the TME and micro-anatomical constraints influence cancer evolution. This situation begins to change with the availability of spatial genomic, transcriptomic and proteomic technologies that enable charting of the growth patterns of distinct subclones and characterisation of the composition and structure of their microenvironments. Connecting these different levels of intratumour heterogeneity is therefore key to understanding how tumours grow as it reveals how tumour subclones interact with their microenvironment and how this may lead to the selection of aggressive traits.

Although there is a broad range of technologies with different advantages and disadvantages, it emerges that cellular resolution is essential for understanding the micro-anatomy of the cancer ecosystem and localising molecular signals to specific cells. Mapping how cells are spatially organised into subclones and neighbourhoods and how different cells interface is a key step for understanding the transcriptional diversity revealed by catalogues of scRNA-seq studies. Another consideration is the field of view. A typical primary tumour is more than 1 cm in diameter at diagnosis; hence, technologies enabling the mapping of entire tumour

sections are desirable. Lastly, enabling spatially resolved genomics with the resolution of single nucleotides is important for mapping the subclonal landscape. Whereas spatial transcriptomic and proteomic approaches are reaching maturity and are available as commercial platforms, spatial genomics lags behind. A reason for this is the limited amount of DNA per cell, compared to the much greater number of copies of RNA and protein, which may be compounded by tissue sections only containing fractions of a nucleus. Even though it is possible to calculate copy number alterations using low-coverage DNA sequencing or even RNA sequencing, it remains challenging to estimate phylogenetic relationships and distances. Developing computational approaches such as integrating mathematical modelling [Gatenbee et al. 2022], mathematical oncology and artificial intelligence-based algorithms may help to overcome these challenges and generate further quantitative insights.

These challenges notwithstanding, the opportunities for spatial cancer genomics are formidable. Charting the tumour ecosystem with clonal resolution as well as cellular and molecular details of the TME will have great value. Doing so across whole tumour sections and cancer types will create insightful atlases that not only provide snapshots of cancer composition but also insights into past tumour evolution. In addition to a more detailed understanding of tumour growth, these atlases have the potential to reveal the mechanism by which the microenvironment is co-opted by tumour cells and how the immune system suppresses tumour proliferation. These processes are already exploited by immune-checkpoint inhibitors, and enhanced molecular and spatial detail may yield further therapeutic targets or facilitate more effective use of existing therapies. Furthermore, the derivation of spatial biomarkers and the detection of diagnostic applications of spatial genomics are likely to inform treatment; this is especially the case now that spatial genomic, transcriptomic, proteomic and metabolomic methods are starting to be used in clinical trials. Such a multimodal approach is required to elucidate the inherent complexity of the TME — the result will be a revolution of histopathology that is genomically and molecularly informed.

The availability of various single-cell and spatially resolved assays and the emerging insights that these approaches offer thus warrant new concerted efforts. Previous initiatives, such as the Cancer Genome Atlas Research Network et al. [2013], the ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium [2020] and International Cancer Genome Consortium et al. [2010], have created valuable community resources detailing the molecular and genomic properties across cancers. It thus seems natural for emerging projects, such as The Human Tumour Atlas Network [Rozenblatt-Rosen et al. 2020], to follow in this spirit in

order to create the next generation of molecularly resolved cancer atlases.

1.7. Aims of this thesis

As highlighted in the preceding section, the spatial aspects of cancer biology hold significant importance in understanding the evolutionary dynamics of cancer. The rapid development of spatial omic technologies in recent years has facilitated the study of cancer biology's spatial dimensions at an unprecedented resolution. The genetic component is particularly vital, as its alterations are not only the fundamental drivers of cancer evolution but also record changes over time. This recording allows for insight into the cancer's history and reasoning at a temporal level. Nonetheless, constructing the genetic maps of tumours with clonal resolution continues to be a formidable challenge. As a result, further efforts must be directed towards developing methods to trace cancer lineages in space. Without this genetic information, the full complexity of cancer's development and behaviour would remain hidden, restricting the opportunities for accurate diagnosis and tailored treatment.

In this thesis, I focus on the development of computational methods to analyse data produced by a promising microscopy-based spatial genomics approach called base-specific *in situ* sequencing (BaSISS) [Lomakin et al. 2022]. This method accurately captures the spatial distribution of a wide range of somatic mutations across large-scale tumour tissue sections. Then I apply designed algorithms to chart clonal maps and deeply investigate two multifocal breast cancers that span the main histological stages of breast cancer progression: ductal carcinoma *in situ*, invasive cancer and lymph node metastasis.

In Chapter 2 I introduce the motivation for the BaSISS experimental protocol and provide a brief foundation in Bayesian statistics necessary to model the experimental data. I then examine the properties of the data and motivate the design of the model, providing a detailed outline of the model's structure and referring to the biological and technical aspects of the system modelled system. I prove the consistency and accuracy of the quantitative nature of the model predictions, drawing on BaSISS technical replicas as well as LCM–WGS targeted sequencing data and illustrate that the model successfully deals with the noise inherent to the data, making it suitable for describing the spatial evolution of cancer clones. The chapter concludes with a discussion of the necessary model assumptions, their biological roots, limitations, and potential future improvements.

Chapter 3 aims to characterise non-genomic modalities of spatial cancer data.

It explores methods for integrating genomic and non-genomic information to discern phenotypic and microenvironmental variations between cancer clones. Specifically, this chapter outlines the properties of generated histological, immunohistochemistry (IHC), and *in situ* sequencing (ISS) data. Given the sparsity of expression data, I employed hierarchical cell markers for cell-type assignment. I introduce a hierarchical logistic regression model for identifying these markers in the reference scRNA-seq atlas. I aggregate multimodal data across manually-defined microregions and formulate statistical models to describe differences in phenotype and TME among cancer clones.

Chapter 4 focuses on applying the previously discussed methodologies to the two cases of multifocal breast cancer. I set the stage by providing a comprehensive introduction to breast cancer, detailing its clinical and biological features while highlighting unresolved questions concerning breast cancer progression. Subsequently, I present the results of BaSISS workflow application and present insights into the structure, nature, and evolution of breast cancer clones. These insights encompass three critical stages of cancer progression: carcinoma *in situ*, invasive carcinoma, and lymph node metastasis.

Chapter 5 summarises the study and provides a broader perspective on the role of spatial genomics methods in cancer research. I discuss the current limitations and potential future solutions for both the BaSISS and the TME characterisation methods used in this study. I then examine the results produced by the BaSISS workflow for breast cancer cases presented in Section 4.2, situating them within the broader context of unresolved questions concerning breast cancer progression. Finally, I conclude the thesis by discussing the utility of spatial genomics methods in oncology, using specific examples from Cancer Research UK's Cancer Grand Challenges. These examples underscore the methods' value in addressing fundamental research questions in the field.

Spatial Genomics: From Noisy Signals to Accurate Clonal Maps

2

Contributions

This chapter is largely based on the supplementary methods and technical results from:

Artem Lomakin, Jessica Svedlund, Carina Strell, Milana Gataric, Artem Shmatko, Gleb Rukhovich, Jun Sung Park, Young Seok Ju, Stefan Dentro, Vitalii Kleshchevnikov, Vasyl Vaskivskyi, Tong Li, Omer Ali Bayraktar, Sarah Pinder, Andrea L Richardson, Sandro Santagata, Peter J Campbell, Hege Russnes, Moritz Gerstung, Mats Nilsson and Lucy R Yates [Nov. 2022]. 'Spatial genomics maps the structure, nature and evolution of cancer clones'. en. In: *Nature* 611.7936, pp. 594–602.

The work I present here is primarily my own contribution. I focused on developing and implementing the core mathematical model for the BaSISS data under the supervision of M.Ger., with valuable inputs from Artem Shmatko and Vitalii Kleshchevnikov. I analysed and interpreted the data, and drafted the original article and figures, again under the supervision of M.Ger. and L.R.Y.. You can find all the code on Github.

Another first author of this paper, J.S., in collaboration with P.J.C., M.N., and L.R.Y., designed the initial study of BaSISS. J.S., M.N., and C.S. conducted the experiments and provided the raw BaSISS data. J.S.P, V.V., T.L., and M.Gat. preprocessed and decoded this data. L.R.Y. performed the LCM cuts, and S.D. conducted the WGS subclonality analyses. I had access to the decoded BaSISS and LCM-WGS data.

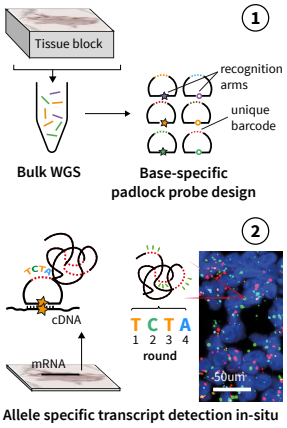
The Introductory Section 2.1.2 and Section 2.1.3, Data Characterisation Section 2.2.1, and Validation Section 2.3 are original. I reworked the rest of the sec-

tions to fit the thesis format, rewording, expanding and illustrating them for clarity. All the main and margin figures are original except Figure 2.2, Figure 2.5, Figure 2.4 and Figure 2.6, which I borrowed from the original paper with stylistic and compositional adjustments.

2.1. Background

2.1.1. The BaSISS protocol

Lineage tracing through somatic mutations can reveal ancestral connections between cancer subclones (Chapter 1). Yet, current methods fall short in preserved human tissue studies [Casasent et al. 2018; Jamal-Hanjani et al. 2017; Jones et al. 2008; Shah et al. 2009; Tarabichi et al. 2021; Yates et al. 2015]. Techniques like laser capture microdissection (LCM) combined with targeted nucleic acid library sequencing, including single-cell sequencing, help to some extent in delineating subclone spatial structures [Casasent et al. 2018; Shen et al. 2000]. However, LCM-based methods fail to capture the full diversity of cancer clones across entire tumour sections. Even the most exhaustive sampling approach would struggle to profile square centimetres at the desired cellular or slightly above cellular resolution. While recent spatial genomics techniques can identify cancer clones based on distinct copy number profiles [Erickson et al. 2022; Zhao et al. 2022], they fall short in detecting point mutations or quantifying mixed clones.



The BaSISS protocol has two essential stages: 1. Bulk WGS identification of clones and mutations, 2. multiplex ISS imaging

Although *in situ* hybridisation [Janiszewska et al. 2015] and mutation-specific padlock probes [Baker et al. 2017; Grundberg et al. 2013; Ke et al. 2013; Larsson et al. 2010] can detect individual mutations *in situ*, their potential is limited by the number of available fluorophores, allowing them to trace only up to 4 mutations at once. Considering the genetic uniqueness of each cancer and its subclones, tracing multiple specific somatic mutations is crucial to decipher spatial and temporal ancestral relationships [Nik-Zainal, Van Loo et al. 2012]. The solution lies in encoding *in situ* sequencing probes combinatorially, allowing them to be captured with a multi-round imaging strategy, thus enabling multiplex base-specific *in situ* sequencing (BaSISS).

The BaSISS protocol uses bulk WGS and z-stacked tissue sections for spatial clone mapping. Full technical details of the protocol are provided in Appendix A. After identifying subclones from bulk WGS, BaSISS employs two main steps. First, it uses padlock probes designed for both mutant and wild-type alleles, marked with a unique 4–5 nucleotide barcode for multiplexing [Ke et al. 2013].

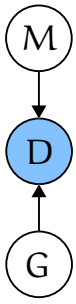
These probes can target any expressed somatic mutation, including rearrangement breakpoints, and can be supplemented with copy number alterations. Next, BaSISS uses a cyclical microscopy approach, akin to gene expression *in situ* sequencing protocol [Ke et al. 2013; Svedlund et al. 2019] (Box 1).

The procedures described, along with initial data processing, produce two primary sets of information. WGS reveals anticipated clonal expansions, identified by point mutations and copy number variations specific to each clone. The spatial component locates alleles identified by WGS in the tissue section. We also obtain a proxy for cell locations by analysing segmented nuclei positions. Nevertheless, the interpretation of subclone locations becomes complicated due to the inherent complexity of the cancer genome, tissue heterogeneity, and potential systematic errors in the technology. The latter poses significant challenges due to the relatively low sensitivity of the BaSISS data (< 1 signal per cell) and off-target hybridisation of the padlock probes (Section 2.2.1 and Figure 2.2, also discussed in Section 2.2.4). These challenges drive the need for a robust statistical approach.

2.1.2. Popular approaches in modelling spatial transcriptomics data

Spatial inference is a common challenge spanning various research fields such as economics, geology, and ecology. Consequently, the field of spatial statistics has become rich in tools and methodologies, often finding applications well beyond their original purpose. For instance, Gaussian process invented solely for geostatistics [Krige 1951], now find use in diverse areas such as optimisation problems, finance, and astronomy. This cross-applicability of tools is indicative of the universal nature of spatial problems. Even in spatial genomics, a relatively recent field, there exists a wide array of mathematical tools that can be applied to solve its unique challenges.

latent variable model *presented as probabilistic graphical model*



Latent (hidden) variables M *and* G *control behaviour of the observed variable* D

While the overarching framework of spatial statistics offers many general tools, each field inevitably cultivates its own preferred methodologies. These preferences often arise from the specific nature of the data and the underlying scientific understanding. In spatial genomics, the uniqueness stems from both the type of data generated and the biological insight that guides the analysis. Often, the task narrows down to mapping and characterising cells, the structural and functional units of the body, using precise markers and observed spatial distributions. Researchers frequently perform these tasks using latent variable modelling. Here, observed spatial distributions $D_{x,y,a}$ of features a (e.g.

expression) in spatial dimensions x, y relate to the spatial distribution $M_{x,y,c}$ of c common factors $G_{c,a}$ (e.g. cell types or expression programmes), as illustrated by:

$$D_{x,y,a} = f(M_{x,y,c}, G_{c,a}) \tag{2.1}$$

The mapping f may be a complex neural network [Dong et al. 2022; Y. Ma et al. 2022] or as simple as a linear operator:

$$D_{x,y,a} = \sum_{c} M_{x,y,c} \cdot G_{c,a} \tag{2.2}$$

However, even in the simplest case of a linear mapping, an essential consideration emerges: if the latent variables remain unconstrained, there is no assurance that the latent representation in $G_{\cdot,a} \in \mathbb{R}^c$ will mirror the actual cell types. Generally, this decomposition is not uniquely identifiable, leading to significant issues for interpretation and numerical estimations.

This necessitates the imposition of constraints, whether applied individually or collectively. Constraints may apply to the spatial component M, utilising tools such as Gaussian process (GP) [Townes et al. 2023], Markov Random Fields [Petukhov et al. 2022], and Conditional Autoregression [Y. Ma et al. 2022; Ni et al. 2022].

Alternatively, constraining the latent variable G is achievable by leveraging the complementary aspects of single cell or bulk genomic data. This includes learning a combined latent space representation from both single cell and spatial data or extracting "cell-type" signatures to guide the spatial model. The latter approach, often referred to as *cell-type deconvolution*[1], correlates closely with the conjecture for BaSISS data. This requires mapping of clones c — cell types defined by specific combinations of known allelic variants a — in tissue space based on these alleles' spatial distribution.

Recent benchmarking within the wide array of spatial transcriptomics models has shown that generative Bayesian models excel in cell type decomposition [Haoyang Li et al. 2023]. Considering additional favourable properties such as interpretability and direct uncertainty estimation, this class of models often becomes an appealing choice for modelling tasks within the field.

[1] In my opinion, a more appropriate name is the alternative term *cell-type decomposition*. While I can see the metaphorical value, it is confusing that the term deconvolution is used on the data where no explicit convolution operations were modelled

2.1.3. Bayesian statistical modelling

In many practical contexts, we find ourselves in situations where the underlying mechanisms governing the system are hidden or only partially observed, and the observations are noisy and limited. Such scarcity of information introduces inherent uncertainty in our understanding of the system. It is within this landscape of uncertainty that the Bayesian understanding of probabilities emerges as a natural tool. Rather than mere frequencies, probabilities are construed as representations of the degree of belief, capturing the epistemic uncertainty about the system parameters. A statistical model, crafted within this Bayesian paradigm, acts as a bridge between the observed data and the underlying phenomena. This model translates our assumptions and prior knowledge into a coherent probabilistic framework. This in turn allows us to make predictions, perform inference, and deeply understand the structure of the data.

Basics of Bayesian inference

Given a set of observations $D = \{t_1, t_2, t_n\}$, a statistical model aims to describe how these observations are generated. One can consider these observations a realisation of a random variable T which is controlled by a set of parameters θ. Therefore, a statistical model could formally be defined by specifying the joint probability distribution $p(D, \theta)$, describing the stochastic relationship between the observations and the controlling parameters. We can rewrite the joined probability using conditional probability, leading us to the Bayes' theorem equation:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} = \frac{p(D|\theta)p(\theta)}{\int p(D|\theta)p(\theta)d\theta} \tag{2.3}$$

Here $p(D|\theta)$ is the likelihood, representing the probability of the observations given the parameters; $p(\theta)$ is the prior distribution over the parameters, reflecting our beliefs about θ before observing the data; $(D|\theta)$ is the posterior distribution after observing the data; and $p(D)$ is the evidence, obtained by marginalising θ out of the joined distribution. It is essentially a normalising constant, ensuring that posterior is a valid probability distribution and integrates to one. Both frequentist and Bayesian models rely on the likelihood $p(D|\theta)$, but they interpret the parameters θ differently. In the frequentist perspective, θ is fixed and its confidence intervals are estimated by considering multiple repeated samples of D. In the Bayesian view, θ varies for each observation set of D and uncertainty in its estimation is expressed as a probability distribution. With this understanding, we

can view statistical inference as an update on our prior belief of the distribution of hidden parameters based on additional evidence.

$$\text{posterior} \propto \text{likelihood} \times \text{prior} \tag{2.4}$$

To demonstrate, let's consider a simple gamma-poisson model of the gene expression, with the gene expression rate as a hidden variable. Consider an observation D_n of observed counts in n cells. I will model the gene expression count for each cell i as a Poisson random variable with the rate parameter λ which has a Gamma prior, where α and β are the hyperparameters shaping our initial belief about expression rate.

$$D_i \sim \text{Poisson}(\lambda) = \frac{\lambda^{D_i} e^{-\lambda}}{D_i!} \tag{2.5}$$

$$\lambda \sim \text{Gamma}(\alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \tag{2.6}$$

Using the fact that the likelihood is independent Poisson distributions and the prior is a Gamma distribution, it is trivial to derive the analytical form of the posterior $p(\lambda|D)$:

$$p(\lambda|D) \propto p(D|\lambda)p(\lambda) \tag{2.7}$$

$$\propto \prod_{i=1}^{N} \frac{\lambda^{D_i} e^{-\lambda}}{D_i!} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \tag{2.8}$$

$$\propto \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{(\sum_{i=1}^{N} D_i)+\alpha-1} e^{-\lambda(n+\beta)} \tag{2.9}$$

$$= \text{Gamma}(\alpha + \sum_{i}^{n} D_i, \beta + n) \tag{2.10}$$

The graphical depiction of the components is shown on Figure 2.1.

Bayesian inference in practice

The process of inference, while appearing simple in principle, presents substantial difficulties in actual computation. A major obstacle lies in the denominator of the Bayesian formulation, the integral often becomes intractable thus rendering even

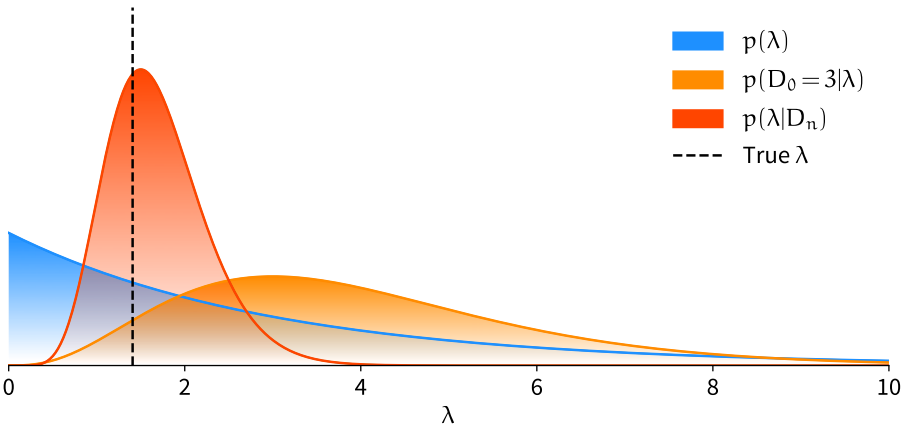


Figure 2.1: Simple gamma-poisson model of cell expression.. Toy mathematical model of gene expression in cells trained on $n = 5$ observations. The gene expression rate λ is a hidden variable, which is modelled as a Gamma random variable with the hyperparameters α and β which is consistent with the week prior belief on low expression. The observed counts D_i are modelled as Poisson random variables with the rate parameter λ. An exact analytical solution for the posterior distribution $p(\lambda|D)$ is tractable (Equation (2.10)) due to the conjugacy of the prior and likelihood.

moderately complex Bayesian models intractable, apart from a few cases with carefully selected *conjugate priors* like shown in the previous example. This issue is a significant constraint in the applicability of Bayesian models, which must be carefully considered in their practical implementation and model formulation.

conjugate priors *are prior distributions that, when combined with a specific likelihood function, result in a posterior distribution with the same functional form.*

The solution to the problems came through a sequence of developments in Marcov chain Monte Carlo (MCMC) methods, including the Metropolis algorithm [A. W. Rosenbluth et al. 1953], Metropolis-Hastings algorithm [Hastings 1970], Gibbs sampling [Geman et al. 1984], and Hamiltonian Monte Carlo (HMC) [Duane et al. 1987]. These methods permit the generation of samples from the posterior probability distribution by forming an reversible Markov chain. The equilibrium distribution of this chain corresponds to the targeted posterior distribution [Roberts et al. 2004]. Consequently, the constructed Markov chain is guaranteed to converge to the true posterior distribution of the model as the number of sampling steps approaches infinity. In practical terms, if the model is well-specified, convergence typically occurs within a reasonable time frame. One can then use samples from the converged chain to estimate necessary summary statistics.

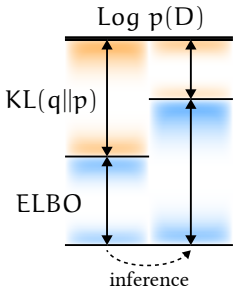
MCMC has revolutionised the field of Bayesian inference, segregating the procedures of model formulation and inference. It must be noted, however, that these methods do not scale with ease when faced with the dimensions and complexity of the model. They often either demand an extensive number of steps or involve calculations that, although fewer in number, are considerably computationally demanding. In scenarios where the analytical posterior is inaccessible, and the computational cost of MCMC is substantial, it is common to use alternative approximation schemes such as variational inference (VI).

Variational inference

Variational inference is framed as an optimisation problem, where the objective is to identify a distribution within a constrained family of tractable distributions that most closely approximates the target distribution [Blei et al. 2017]. Mathematically, the problem of variational inference can be formulated as finding the distribution $q(\theta)$ of the variational family Ω that is following the true posterior $p(\theta|D)$ as closely as possible. The common distance measure that is used is the Kullback-Leibler (KL) divergence, thus the goal is:

$$q(\theta) = \underset{q \in \Omega}{\operatorname{argmin}} \, \mathrm{KL}(q(\theta) \| p(\theta|D)) \tag{2.11}$$

After expanding KL divergence and rewriting conditional probability:



Evidence lower bound (ELBO) is directly linked with KL *divergence and evidence* $\log p(D)$. *Thus when increasing ELBO,* KL *divergence inevitably decreases.*

$$q(\theta) = \underset{q \in \Omega}{\operatorname{argmin}} \int q(\theta) \log \frac{q(\theta)p(D)}{p(\theta, D)} d\theta \tag{2.12}$$

$$= \underset{q \in \Omega}{\operatorname{argmin}} \int q(\theta)[\log q(\theta) + \log p(D) - \log p(\theta, D)] d\theta \tag{2.13}$$

$$= \underset{q \in \Omega}{\operatorname{argmin}} \, \mathbb{E}_q[\log q(\theta) - \log p(\theta, D)] + \log p(D) \tag{2.14}$$

Notice that $\log p(x)$ is constant with respect to q, so we can ignore it when minimising KL divergence. Now we can define the optimisation objective, which is the inverse expectation term, referred as the evidence lower bound (ELBO). Hence, the original problem of minimising KL divergence between approximation distribution and posterior becomes equivalent to the maximisation of ELBO.

One significant development in VI is automatic differentiation variational inference (ADVI) [Kucukelbir et al. 2016]. ADVI generalises the approach by employing the normal family of distributions as the guide, thereby enabling the ap-

plication of gradient-based optimisation techniques, via reparameterization trick.

In ADVI, the variational distribution is assumed to be multivariate normal with an appropriate transformation.

$$q(\theta) = f(\mathcal{N}(\mu, \Sigma)) \tag{2.15}$$

reparameterization trick *allows for the sampling of a variable* $Z \sim \mathcal{N}(\mu, \sigma^2)$ *to be reformulated as* $Z = \mu + \sigma \cdot \epsilon$, *where* $\epsilon \sim \mathcal{N}(0, 1)$. *This transformation makes it possible to compute gradients with respect to* μ *and* σ.

Here, μ and σ are the optimised mean and covariance parameters, and f is a transformation function that maps a normal distribution to the desired target distribution. The transformation function is chosen such that the support of the transformed distribution is the same as the support of the target distribution. For example, if the target distribution is a Gamma distribution, then the transformation function would be the Exponential function.

In practice, additional constraints are imposed on the covariance matrix, considering it either low-rank or simply diagonal (mean-field) [Blei et al. 2017]. Although these assumptions may overlook a probable covariance structure between model parameters, and frequently underestimate the variance, they dramatically decrease the number of parameters that are optimised. This makes the inference of even larger models feasible.

Probabilistic programming

In addition to advancements in inference algorithms, contemporary machine learning frameworks like Theano [The Theano Development Team et al. 2016][2], PyTorch [Paszke et al. 2019], TensorFlow [Abadi et al. 2016], and JAX [Bradbury et al. 2018], have notably improved the efficiency of Bayesian inference. This is particularly true for methods that utilise automatic differentiation, such as the HMC family techniques and ADVI. Probabilistic frameworks, including PyMC [Salvatier et al. 2016], Pyro [Bingham et al. 2018], and NumPyro [Phan et al. 2019], are simplifying the Bayesian model development process. By managing most of the underlying calculations, these frameworks allow researchers to concentrate more on conceptual model development and spend less time on numerical details.

BaSISS profiles extensive areas of tissue, encompassing hundreds of thousands of cells. The underlying processes that generate the final data are complex. Therefore, defining a statistical approach becomes unfeasible without the utilisation of cutting-edge inference approaches and frameworks capable of handling this com-

[2] A discontinued python framework that I had a misfortune to use as a backend for model optimisation defined in PyMC package. Currently a fork is maintained under the name Aesara

plexity.

2.2. A Bayesian model to localise cancer clones in tissue

2.2.1. BaSISS data characterisation

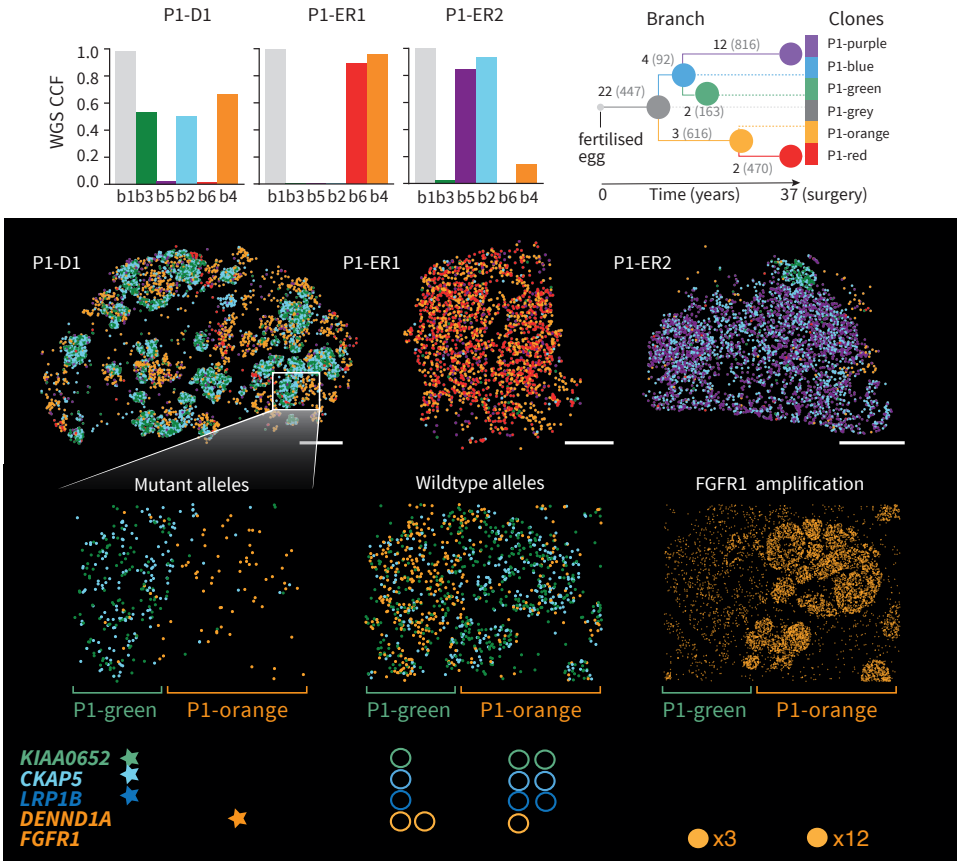


Figure 2.2: Raw BaSISS data for three samples from P1 (P1-ER1, P1-ER2, and P1-D1). Selected targets are coloured by the clonal origin of the mutated allelic variant. WGS estimated cancer cell fraction and clonal phylogeny are shown on the top. The bottom panel presents the link between clone genotypes and BaSISS signal density. Scale bar = 2.5 mm

Before diving into the model design, let's first examine the raw BaSISS data for three samples from the patient P1 (P1-ER1, P1-ER2, and P1-D1). Detailed descriptions of these samples and the associated experimental procedures can be found in Section 4.2.2 and Appendix A. In brief, the multi-region WGS identified six clonal

expansions, and 51 padlock probes were designed to track them on tissue slides targeting mutant and wild-type allelic variants (Appendix A.5.3 and Table A.1). In total, these three samples contain approximately 1,000,000 cells, spanning 245 mm^2 of tissue (Table D.2), and roughly 1,000,000 detected BaSISS signals, 25% of which target mutant variants.

If one is to colour allelic variants by their clonal origin, it is easy to note that the data has hierarchical structure Figure 2.2. For example private mutations of P1-green and P1-blue clones tend to co-localise, while avoiding P1-orange clone private mutations. This is consistent with the clonal phylogeny, where P1-green is a descendent of P1-blue, and therefore possesses all P1-blue mutations, while P1-orange clone is a separate lineage. When it comes to the corresponding wild-type alleles the situation is reversed, in the suspected location of P1-green clone, wild-type variants of the genes where it has a mutation is depleted, which makes sense as the copy number of the wild-type variant is decreased once mutation happens. An even more obvious relation between copy number and signal density is observed for the copy number variable locus of *FGFR1*. Since P1-orange clone possesses a greater copy number gain of 12 in this locus, the signal density is increased in the corresponding region Figure 2.2.

These observations motivate two key model design decisions. First of all, it is obvious that I won't be able to extract single cell level information from the data, as the signal density is too low. Therefore, I will be working with aggregated data, where the unit of observation is a region, not a cell. Cells inside these regions can belong to different clonal lineages and each should contribute to the observed signal count. Secondly, there is a clear relation between the copy number and signal density, which suggests that I should incorporate copy number information into the model.

2.2.2. Non-negative matrix factorisation

As discussed previously, the BaSISS protocol involves recording a series of fluorescent spots, each of which corresponds to a targeted allele. These spots are decoded into a class of different barcodes A, resulting in a table of tuples (a_i, x_i, y_i), where a_i represents the allele of spot i, and x_i and y_i denote its respective two-dimensional coordinates on a grid.

This grid can be visualised as a three-dimensional array $\mathbf{D} \in \mathbb{N}^{|a|\times|x|\times|y|}$, where a refers to the specific allele, and x and y correspond to coarse-grained coordinates on a regular grid of dimensions $|x| \times |y|$. The grid size was selected to be 108.8 μ m, a choice that takes into consideration a balance between data

sparsity, precision, and computational cost.

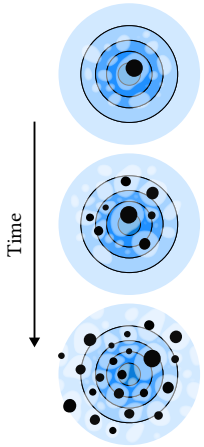
The essential concept behind this approach is that the expected number of BaSISS signals can be decomposed into maps of $|s|$ distinct clones s, each represented as $\mathbf{M} \in \mathbb{R}^{|s|\times|x|\times|y|}$, and a genotype $\mathbf{G} \in \mathbb{N}^{|a|\times|s|}$,

$$\mathbb{E}[\mathbf{D}] \approx \mathbf{G} \times \mathbf{M} = \sum_{s} \mathbf{G}_{a,s}\mathbf{M}_{s,x,y} \tag{2.16}$$

The genotype matrix $\mathbf{G}$ encapsulates the information about the number of each allelic copy present in each clone s. Essentially, $\mathbf{G}$ serves as a matrix representation of the underlying phylogenetic tree, providing a detailed guide to the allelic configuration at each branch of the tree.

The maps $\mathbf{M}$, on the other hand, offer insight into the relative prevalence of each clone in a specific area of the grid. The maps are modelled as non-negative matrices, so the number of cells in a given region cannot be negative. The non-negativity constraint is biologically relevant, as it is impossible for a clone to have a negative prevalence in a given region.

Since all three matrices are non-negative, the problem can be categorised as a *non-negative matrix factorisation* problem. It is a known characteristic of such problems that they typically do not have unique solutions. Therefore, additional constraints, guided by biological reasoning, are required to ensure the identifiability of the solution. While the matrix $\mathbf{G}$ possesses a clear biological interpretation and is essentially fixed, the spatial nature of matrix $\mathbf{M}$, being a latent variable, must be carefully defined.



Cancer growth can be effectively modelled using the RBF kernel, as the cohesion in many tumours causes descendant cells to stay close together, mirroring the localised similarity patterns that this kernel captures.

2.2.3. Latent Gaussian process

Since the matrix $\mathbf{M}$ is designed to represent a spatially continuous map, we can utilise the inherent assumptions about spatial locality present within the data. Specifically, it is reasonable to anticipate that points that are spatially closer to one another should display higher similarity than those further apart. This expectation not only arises from a general consideration of spatial smoothness but also finds roots in biological rationale. Considering the process of cancer growth, for example, dividing cells are typically situated near their parent cells, which further accentuates local similarity. This concept of spatial locality can be mathematically described using a Radial Basis kernel function:

$$K_x(x, x') = \exp\left[-\frac{(x - x')^2}{2l^2}\right] \tag{2.17}$$

Here, x and x' denote spatial coordinates, and l is a length scale hyperparameter that governs the smoothness.

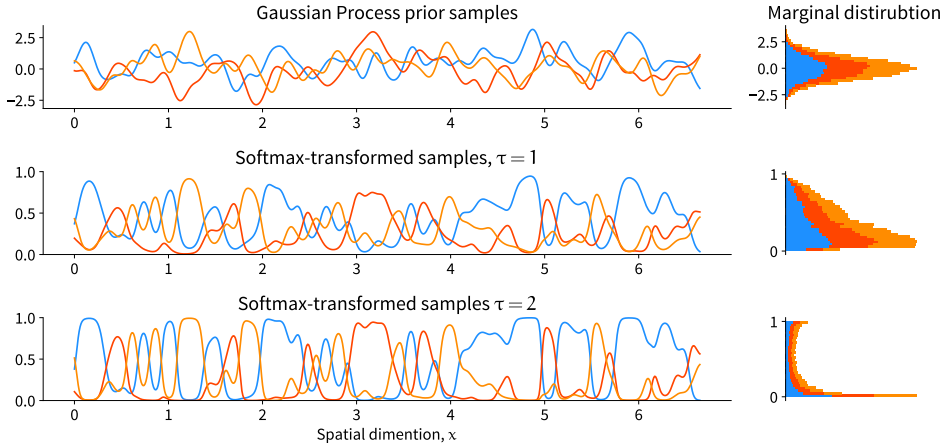


Figure 2.3: Samples from a softmax transformed GP under different τ. Samples from a softmax transformed GP under different τ. The GP is defined by a squared exponential kernel with length scale $l = 0.1$. The samples are transformed using a softmax transformation with different temperatures τ. The temperature hyperparameter controls the level of intermixing between the clones

We can naturally express a generative process constrained by the kernel function through the Gaussian process (GP). The GP represents a stochastic process where any finite set of random variables exhibits a joint Gaussian distribution. Put simply, a GP serves as a distribution over functions which shape is constrained by the kernel.

The GP relies on two main components for its specification: a mean function $m(x)$ and a covariance function $K(x, x')$. Since I am dealing with a two-dimensional spatial map, where x and y dimensions are independent, the covariance function is defined as the product of two one-dimensional kernels, $K_{xy} = K_x \times K_y$. As there is no prior knowledge about the expected prevalence of clones in the region, the mean function is set to zero:

$$m_s(x, y) \sim \mathcal{GP}(0, K_x \otimes K_y) \tag{2.18}$$

The GP is a distribution over functions that belong to the $\mathbb{R}$ domain. However, I want $\mathbf{M}$ to be non-negative, and even more specifically, to be a probability simplex. To achieve this, I apply a *softmax transformation* to the GP output, which maps the real-valued function to a probability simplex. The softmax transforma-

tion is defined as follows:

$$\mathbf{M} = \left[\frac{e^{\tau m_0}}{1 + \sum_{i=0}^{|s|-1} e^{\tau m_i}}, \ldots, \frac{e^{\tau m_{|s|-1}}}{1 + \sum_{i=0}^{|s|-1} e^{\tau m_i}} \right] \tag{2.19}$$

Here, τ is a temperature hyperparameter that controls the prior expectation of clone intermixing. By setting high τ values, we express our belief that clones are likely to be spatially separated, while higher values of τ indicate a higher degree of intermixing as shown on Figure 2.3. Since I didn't have a strong expectation about the degree of intermixing, I set $\tau = 2$ for the case with 6 clones, which corresponds to a weak belief that clones tend to intermix.

The use of GP as a prior for the spatial maps $\mathbf{M}$ is a convenient way to deal with the sparsity of the data observed in the BaSISS data. It allows to propagate information over the spatial dimensions in a non-parametric way, stabilising the solution and reducing the impact of sampling fluctuations. Nevertheless, the model requires further explicit parametrisation of the known sources of variation.

non-parametric *Even though GP include kernel hyperparameters, the term "non-parametric" here refers to the fact that the GP doesn't assume a fixed form for the underlying function it's modelling. The model can adapt to the underlying structure in the data and can represent an infinite variety of functional forms.*

2.2.4. Sources of variation

I model the expected number of BaSISS signals, denoted by $\mathbb{E}[\mathbf{D}]$, within a specific region. This calculation involves two components: the genotype matrix $\mathbf{G}$ and the spatial map $\mathbf{M}$. The matrix $\mathbf{G}$ encodes integer values representing the allelic copy number within each cell of a particular clone, while the spatial map $\mathbf{M}$ shows the relative prevalence of each clone in the region.

The product of $\mathbf{G}$ and $\mathbf{M}$ provides a value that one might interpret as the average allelic copy number in the given region. However, this interpretation differs significantly from the "expected number of BaSISS signals" which can be lengthy described as the expected number of fluorophores observed as a result of sequencing by ligation of the amplified padlock probes hybridised with cDNA reads synthesised from the mRNA expressed by cells in the region. Hence, in order to rescale the product of $\mathbf{G}$ and $\mathbf{M}$ to the expected number of BaSISS signals, I need to account for the following sources of variation:

1. Cellular density ν
2. Differential probe specificity ι
3. Allelic confusion τ
4. Clone expression variations γ

5. Homogeneous and inhomogeneous background adjustments β

6. Overdispersed sampling fluctuations α

Accounting for these, the equation for the expected number of BaSISS signals Equation (2.16) becomes:

$$\mathbf{M}_{a,x,y} = \underbrace{\nu_{x,y}}_{\text{cell density}} \cdot \overbrace{\iota_a}^{\text{detection rate}} \cdot \underbrace{\sum_{a'} \tau_{a,a'}}_{\text{probe confusion}} \overbrace{\sum_{s} \gamma_{s,a}}^{\text{clone-specific expression}} \underbrace{\mathbf{G}_{a,s}\mathbf{M}_{s,x,y}}_{\text{clone contribution}} + \overbrace{\beta_a}^{\text{background}} \tag{2.20}$$

Let's discuss each of these components in more detail.

Cellular density

The cellular density $\nu_{x,y}$ directly influences the number of observed signals in a given region. While this density represents the expected number of cells, directly observing it poses a challenge. Some tools attempt to infer cellular density from transcriptional signals alone [Kleshchevnikov et al. 2022]. However, this method is unreliable due to confounding expression specific factors.

confounding expression
Same signal level could be explained by higher global expression, or more dense cell populations.

Fortunately, the number of nuclei in the region provides necessary information to disentangle the effect of cellular density from the other sources of variation. We obtain this data during the BaSISS signal acquisition process. Using a neural network model, I segment the DAPI stained image of nuclei, saving the nuclei counts in a two-dimensional array $\mathbf{N} \in \mathbb{N}^{|x|\times|y|}$. These counts may not always accurately reflect cell densities. This inaccuracy can result from segmentation errors or because the slide captures only a fraction of the cells. As a result, I model the nuclei counts $\mathbb{N}_{x,y}$ as Poisson distributed values. Their mean, $\nu_{x,y}$, represents cell densities and follows a weak Gamma prior.

$$\mathbf{N}_{x,y} \sim \text{Poisson}(\nu_{x,y}) \tag{2.21}$$

$$\nu_{x,y} \sim \text{Gamma}(\mu, \sigma) \tag{2.22}$$

The prior parameters μ and σ are set to a weekly informative distribution center around the expected number of cells in the region.

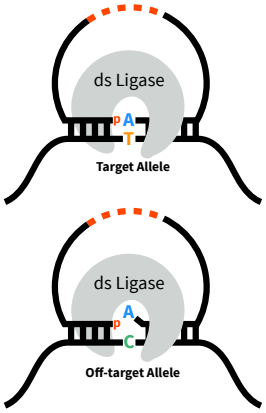
Differential probe specificity

BaSISS ability to detect allelic variants relies on a multistage process of reverse transcription, padlock probes annealing, amplification and fluorophores detection [Svedlund et al. 2019]. As it is hard to model each of the steps explicitly, I model the overall detection rate $\iota \in \mathbb{R}_{+}^{|a|}$ for each allele a as a Gamma distributed random variable.

$$\iota_a \sim \text{Gamma}(\mu, \sigma) \tag{2.23}$$

The prior parameters μ and σ are set to a weekly informative distribution with the probability density shifted towards zero, as I do not know the average probe specific detection rate, but it is likely to be low given that I observe < 1 BaSISS signal per cell.

Allelic confusion



Even with the mismatch between padlock-probe and target sequence, DNA Ligase is capable of ligating the break.

The difference between wild-type and mutated allelic variants often amounts to a single nucleotide. This elevates the chances for the padlock probe to anneal to the off-target allelic variant, resulting in allelic confusion. Though DNA ligases are optimised for fidelity, they can make mistakes, particularly under suboptimal conditions [Lohman et al. 2016]. This mistake-prone behaviour particularly contaminates mutated allele reads. Since most cells contain at least two wild-type copies, this creates spurious noise, as seen in Figure 2.2 (red dots in P1-D1).

To model this effect, I design a sparse transition matrix $\tau \in \mathbb{R}_{(0,1)}^{|a|,|a|}$, populating it with transition probabilities $\{\tau_{a,a'}\}$ for each allele. I impose a strong regularising Beta prior on $\tau_{a,a'}$, shifting the probability density close to 1 and preventing it from dropping below 0.75. This restriction avoids non-identifiability and is a reasonable constraint, given that this effect is selected against during protocol optimisation.

$$\tau = \left(\begin{array}{c|cc} & \dots & \text{wt} \quad \text{mut} \quad \dots \\ \hline \vdots & & \\ \text{wt} & & \tau_{11} \quad 1-\tau_{11} \\ \text{mut} & & 1-\tau_{22} \quad \tau_{22} \\ \vdots & & \end{array}\right), \tau_{a,a'} \sim 1 - \text{Beta}(\alpha, \beta)/4 \tag{2.24}$$

I set the values corresponding to different loci to zero, as off-target probe hy-

bridisation is unlikely to result in a signal.

Clone specific allelic variations

The detection rate matrix, ι, registers the mean clonal expression level. However, we can reasonably assume that different clones exhibit some level of variability in gene expression. To model this deviation, I use a matrix $\gamma \in \mathbb{R}_+^{|s|\times|a|}$, with a LogNormal distributed prior such that $\mathbb{E}[\gamma] = 1$.

$$\gamma_{s,a} \sim \text{LogNormal}(\mu, \sigma) \tag{2.25}$$

Regularising this prior is crucial to ensure that most of the probability density surrounds 1 by choosing a small σ. Since the matrix γ has the same shape as the genotype matrix $\mathbf{G}$, failing to control this parameter could lead to non-identifiability.

Homogeneous and inhomogeneous background adjustments

Raw BaSISS data are derived from biological images that include additional sources of pixel-level variability. The first source is a homogeneous additive shift $\beta \in \mathbb{R}_+^{|a|}$, which globally increases the expected value of a particular probe detection on the slide. This is consistent with a potential global background autofluorescence which differs in the wavelength, thus affecting each probe independently, as they are combinatorially encoded by fluorophores. In combination with the failed probe detection, this effect can lead to a global shift in the signal density. I model this effect as a Gamma distributed random variable.

$$\beta_a \sim \text{Gamma}(\mu = 0.5, \sigma = 1) \tag{2.26}$$

The second source is an inhomogeneous background shift that adjusts for the variability in local base signal detection. Such factor should help the model to explain changes in signal composition that are not explainable by the linear combination of a known clonal genotypes. In practice, besides unknown clones, this factor tend to help with the technical noise, when some of the tissue areas are not properly registered, or in the necrotic areas, where nuclei count is high, but the BaSISS signal is low.

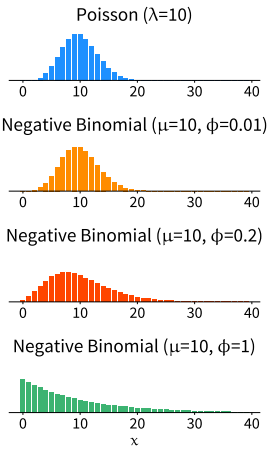
Modelling the variance of this type without imposing harsh regularisation presents a challenge due to its flexibility. To represent the inhomogeneous background shift, I introduce additional pseudo-clones, denoted by ψ, with

corresponding relative prevalence $\mathbf{M}$ and pseudo-genotype $\Gamma \in \mathbb{R}_{+}^{|p|\times|a|}$. The prevalence in this context mirrors that modelled for the actual clones. The pseudo-genotype is inferred from a Beta prior, which is manipulated to stretch the support to match the median copy number k among the loci. Although I use $|\psi| = 1$, this value can be increased to allow for a more flexible background shift.

$$\Gamma_{p,a} \sim \text{Beta}(\alpha, \beta) \times k \tag{2.27}$$

The Beta distribution parameters are chosen to shift the probability density towards zero, which should prevent the pseudo-genotype from being overestimated and overused by the model.

Sampling fluctuations



Negative Binomial is equivalent to the Poisson distribution with a Gamma prior on the mean parameter λ.

I model BaSISS counts as being Negative Binomial distributed given an unobserved mean detection level $\mathbf{M}_{x,y,a}$ and over-dispersion parameter α_a which accounts for unexplained variance:

$$\mathbf{D}_{x,y,a} \sim \text{NB}(\mu_{x,y,a}, \alpha_a) \tag{2.28}$$

This is a common choice of the likelihood function [Kleshchevnikov et al. 2022; Townes et al. 2023], as it is more flexible than Poisson, which assumes that the variance is equal to the mean.

Over-dispersion α_a is sampled from Gamma distribution where distribution density is shifted towards lover value, making the likelihood Poisson like if no over-dispersion is observed.

$$\alpha_a \sim \text{Gamma}(\mu, \sigma) \tag{2.29}$$

This concludes main part of the model definition. The full model is conceptually visualised in Figure 2.4.

2.2.5. Regularisation with auxiliary constraints

Depending on the experimental setup, we may obtain auxiliary data containing information about the state of the system we are modelling. In the spirit of Bayesian inference, propagating this information through the model should reduce uncertainty in the posterior estimation of the parameters. In this section, I describe two types of auxiliary data likely to be generated alongside the BaSISS experiment, which can be used to regularise the model.

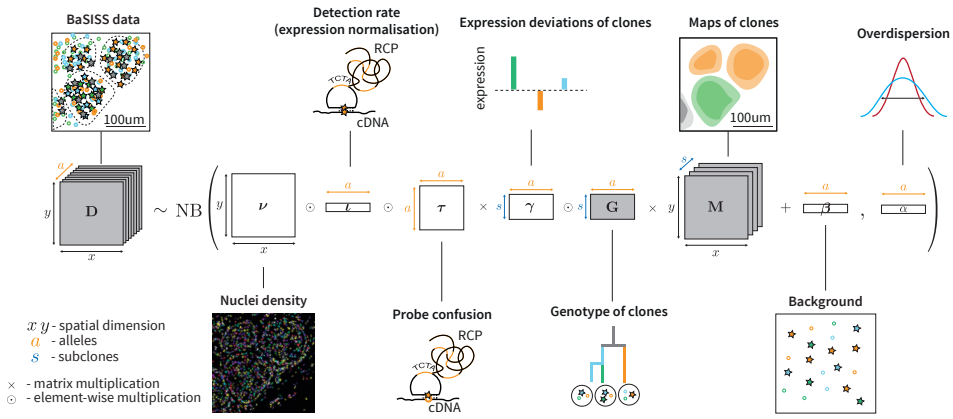


Figure 2.4: Mathematical modelling BaSISS data. Mathematical model for generating quantitative clone maps. The essential idea is that the BaSISS signals count matrix $\mathbf{D}$ is decomposed into maps of clones $\mathbf{M}$ each with a distinct genotype $\mathbf{G}$ (grey shading), accounting for multiple sources of variability.

Bulk variant allele frequency information

WGS forms the initial step of BaSISS, and is vital for inferring clone composition and identifying potential targets for probe design. This step is carried out on the tissue sections adjacent to the block used for BaSISS, as illustrated in Figure 4.5. Since the neighbouring tissue sections are likely to have similar, though not identical, clone composition, we can use the WGS data to regularise the global clonal composition inferred by the model.

WGS directly measures the number of reads for each allelic variant, so the relative variant allele frequency (VAF) for each locus should remain the same across tissue sections with identical relative clone composition. I apply this constraint to the model by constructing an auxiliary variable $\mathbf{VAF}^{\text{BaSISS}}$ of each mutant variant m on the slide. This variable is formed by the summation of $\mathbf{M} \times \mathbf{G}$ product over spatial dimensions and represents the bulk inferred VAF of the slide:

$$\mathbf{VAF}^{\text{BaSISS}} = \sum_{x,y} \mathbf{M} \times \frac{\mathbf{G}^{\text{mut}}}{\mathbf{G}^{\text{wt}} + \mathbf{G}^{\text{mut}}} \tag{2.30}$$

Next, I construct an auxiliary likelihood of $\mathbf{VAF}^{\text{BaSISS}}$ with α and β parameters proportional to the number of mutated and wild-type reads found in the WGS experiment. I incorporate parameter u to account for uncertainty in WGS data since it originates from a proximal slide that is not exactly the same as the one

used in the BaSISS experiment:

$$\mathrm{VAF}_m^{\mathrm{BaSISS}} \sim \mathrm{Beta}(\alpha = \mathrm{WGS}_m^{\mathrm{mut}}/u + 1, \alpha = \mathrm{WGS}_m^{\mathrm{wt}}/u + 1) \tag{2.31}$$

It is important to note that this likelihood contains exactly $|m|$ mutant alleles terms, roughly half the number of alleles $|a|$. In contrast, the main model likelihood Equation (2.20) includes $|x| \times |y| \times |a|$ terms, several orders of magnitude higher. This discrepancy means that the main model likelihood will predominantly influence the posterior estimation. Thus, I multiply the log-likelihood of the auxiliary constraint by a factor of $|x| \times |y|$ when formulating the ELBO.

Immunohistochemistry based cell type counts

Another potential auxiliary data source stems from IHC staining of the tissue section. IHC staining, a standard technique in pathology, serves to identify specific cell types. Within the scope of the BaSISS experiment, IHC staining can be employed to characterise the spatial microenvironment of the tissue section. As the cancer cells' staining by IHC is known, this information can be used to regularise the model.

In our study, I focused on working with CD45 IHC staining, which pinpoints the location of immune cells. As none of the breast cancer cells should be stained by CD45, we can establish a connection between the fraction of stained cells in the region and the cancer cell fraction. This connection is made through the construction of an auxiliary variable for cell fraction, denoted as **CellFrac**, obtained by summing the clone map matrix $\mathbf{M}$ over clones of the respective types s^- (including cancer cells and normal genotypes of non-immune cells) or s^+ (covering immune cells with normal genotypes).

parameters α ***and*** β *The beta-binomial distribution models the number of successes in a fixed number of trials, where the probability of success follows a beta distribution. The parameters* α *and* β *can be interpreted as "pseudo-counts," with* α *representing 1 + number of prior successes, and* β *1 + number of prior failures.*

$$\mathbf{CellFrac} = \frac{\sum_{s \in s^-} \mathbf{M}}{\sum_{s \in s^+} \mathbf{M}} \tag{2.32}$$

Subsequently, I formulated an auxiliary likelihood of **CellFrac** using the Beta distribution, where the parameters α and β correspond to the cell counts of IHC- and IHC+ stained nuclei on the slide.

$$\mathbf{CellFrac}_{x,y} \sim \mathrm{Beta}(\alpha = \mathrm{IHC}^-_{x,y} + 1, \alpha = \mathrm{IHC}^+_{x,y} + 1) \tag{2.33}$$

This form of regularisation proves especially beneficial in samples with very low cancer cell fraction, such as in lymph node metastasis, where the majority of cells are immune cells. In these instances, the model might mistakenly inter-

pret noise in areas with high normal cell density as cancer cell signals, thereby overestimating the local CCF.

2.2.6. Multi-sample extensions

When conducting a BaSISS experiment on multiple tissue sections belonging to the same cancer patient, and when these sections are acquired and prepared in a batched manner, it becomes possible to share many of the model's parameters between the slides. Specifically, the shared parameters include the genotype matrix $\mathbf{G}$ (since they pertain to the same cancer case), clone-specific allelic expression variation γ (as clones are likely to exhibit similar expression patterns), the probe confusion transition matrix τ, and the expected probe detection rate ι (since the same genes are targeted, and the samples are prepared under identical conditions). The remaining parameters are specific to each of k individual slides.

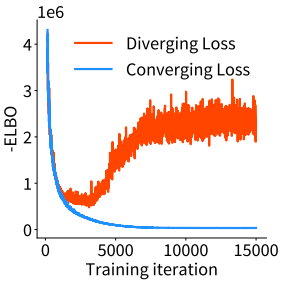
To address potential variations in sample preparation, I multiply the mean probe detection rate by a slide-specific probe deviation matrix $\eta \in \mathbb{R}_+^{|k|\times|a|}$. This matrix has a log-normally distributed prior, with $\mathbb{E}[\eta] = 1$ and a variance of 0.05.

$$\eta_{k,a} \sim \mathrm{LogN}(\mu, \sigma) \tag{2.34}$$

This extension especially pays off when certain clones are present only in one of the slides. It prevents the model from overfitting to the noise within the data.

2.2.7. Inference

Variational inference is used to approximate the posterior, specifically employing the mean-field version of ADVI as found in the PyMC package. As explained in Section 2.1.3, ADVI optimisation approximates the posterior distribution over unknown parameters using an appropriately transformed multivariate normal distribution. This transformation allows the exploitation of the automatic differentiation framework. Since estimating the covariance between parameters proved to be excessively costly, the optimisation is performed only on the mean and diagonal elements of the covariance matrix of the variational distribution.



Numerical instabilities could lead to errors in gradient estimations and an erratic behaviour of the Loss function.

Training is halted when the ELBO no longer rises, typically after at least 15,000 iterations using the ADAM stochastic optimiser [Kingma et al. 2014] with a learning rate of 0.01. For subsequent analysis, the marginalised variational posterior of the parameters is employed. My experiments demonstrate that multiple initialisations may be necessary to reach an optimal solution. This requirement is

primarily driven by the stochastic selection of initial values but might also be affected by numerical instabilities occurring during inference with PyMC.

2.3. Validation

Validating spatial genomics models presents a formidable challenge. Simulation studies are of limited use, as they fail to capture the complexity of real data. Furthermore, the absence of ground truth data precludes a direct comparison of model predictions with actual values. Most validation methods, therefore, rely on indirect strategies. These include assessing correlations between the distributions of conservative marker genes and inferred cell types, or evaluating the stability of the solution on technical or simulated replicates [Haoyang Li et al. 2023]. For the BaSISS experiment, we generated a technical replicate to gauge the stability of both experimental and computational aspects of our approach.

Defining cell-type *The categorisation of cell types is often ambiguous because of transcriptional plasticity. While broader cell types may exhibit stable markers, specialised or niche cell types tend to be more transient and elusive, leading to false overclustering [Grabski et al. 2022].*

Defining cell-type often poses an ambiguous task. However, the BaSISS model seeks to map clones defined by a set of point mutations, offering a robust and often binary definition (either present or not). Consequently, it becomes feasible to generate near ground truth data using the LCM WGS approach [Shen et al. 2000]. In this method, the tissue undergoes microdissection into regions of interest, followed by DNA extraction and sequencing. Though expensive and time-consuming, this strategy yields high-resolution allelic frequency data, allowing for an assessment of model prediction accuracy.

Technical replication and LCM-WGS experiments were carried out on cancer case P1, briefly introduced in Section 4.2.2 and described in more detail in Tables D.1 and D.2.

Consistency of the BaSISS data

In the initial stages of the BaSISS experimental protocol optimisation, J.S., M.N. and C.S. produced two consecutive slide technical replicates over several years. These replicates were created using different probe designs, imaging platforms, and tissue preparation strategies, as detailed in Appendix A.5. This sequence offers an invaluable resource for evaluating the consistency of the experimental protocol, particularly concerning signal distribution.

Figure 2.5 displays the correlations between the VAF distributions of the two technical replicates. Despite the variations in experimental conditions, the BaSISS-derived VAFs demonstrate strong correlation across replicate experiments on serial tissue sections (R = 0.76–0.93, Pearson's). This result underscores

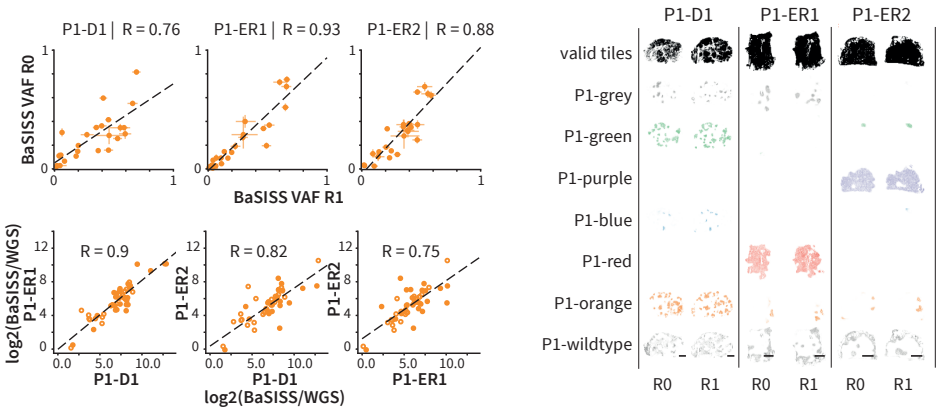


Figure 2.5: Technical consistency of the BaSISS protocol. ***Top Left*** part of the plot shows the Person's correlation between the VAF distributions of the two technical replicates - R0 and R1. Data are presented as mean estimates and 95% HPDI. ***Bottom Left*** confirms the correlation between WGS read normalised BaSISS mutant (solid) and wild-type (hole) allele counts. This systematically confirms the correlation between BaSISS signal density and copy number alteration (CNA) level. ***Right*** part of the plot shows the fields $\mathbf{M}$ inferred by the model for each of the two replicates. The fields are visually similar, considering tissue distortion and distance on z-stack. Scale bar = 2.5 mm

the BaSISS protocol's robustness against changes in experimental circumstances.

I have previously addressed the visual correlation between BaSISS signal density and copy number alteration (CNA) level in Section 2.2.1, citing it as a motivation for the model design. Nonetheless, it is crucial to formally evaluate whether this correlation holds generally across loci. We possess WGS data for the samples, providing us with bulk WGS estimates. Although the precise clonal composition remains unknown, we can reasonably presume that normalising BaSISS signals by WGS-detected reads should account for the compositional differences between samples. As depicted on the bottom-left subplot of Figure 2.5, this assumption holds true, with $\log_2$(BaSISS/WGS) values showing reasonable correlation (R = 0.75–0.9, Pearson's).[3] This finding supports the assumption that BaSISS signal density correlates with the number of DNA copies in the sample.

Consistency of the model predictions

The replication experiment mentioned earlier serves to assess the stability of the model's predictions. As the probes are engineered to target identical mutations,

[3] Apparently, in the original publication I accidentally reported this equation incorrectly as VAFs instead of the raw signal counts.

the model ought to generate comparable clonal maps for both replicates. However, certain variables, such as the probe detection rate, denoted as ι, may vary due to differences in probe barcodes and experimental conditions. When run with identical hyperparameters on the two replicates, the model yields reasonably similar clonal maps, accounting for the distance in the z-stack between them (refer to Figure 2.5).

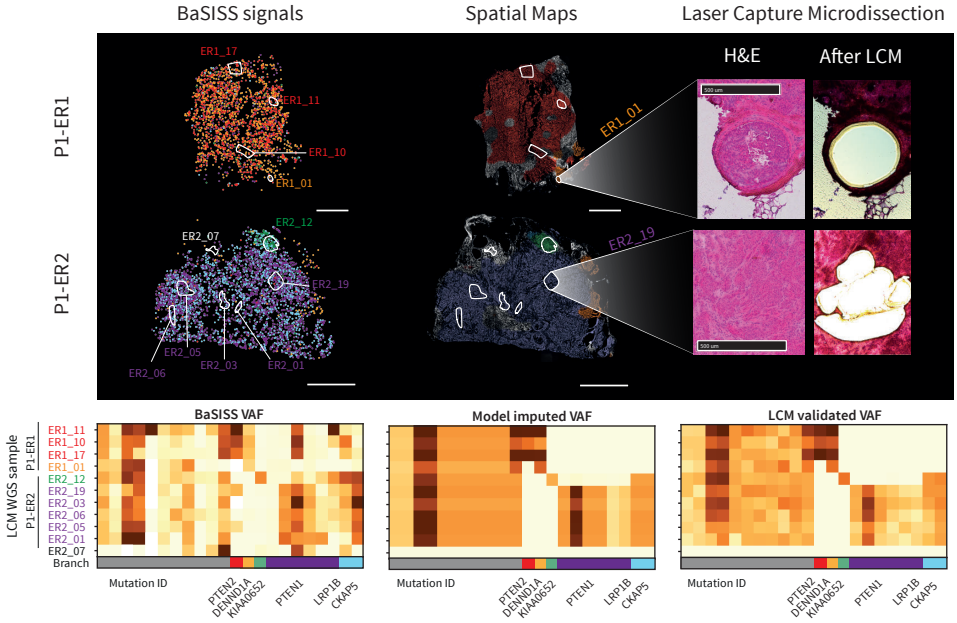


Figure 2.6: LCM-WGS validation of the BaSISS model prediction. ***Top*** part of the plot shows the Person's correlation between the VAF distributions of the two technical replicates - R0 and R1. Data are presented as mean. Scale bar = 2.5 mm. ***Bottom*** shows VAF values for the validated regions: BaSISS signal, which appears noisy, BaSISS model imputed VAF highly concordant with LCM-WGS estimates.

Validation of model predictions with LCM-WGS

Before using the aforementioned model in practice, we must validate the accuracy of its detailed predictions of clonal compositions. Characterising the entire tissue slide is unfeasible due to its laborious nature, but we can select a small number of regions for LCM-WGS to confirm the model's predictions. For this purpose, we performed LCM-WGS on the residual tissue blocks.

It is vital to acknowledge that the tissues used were not adjacent slides but were

located at a significant distance along the z-stack. This distance represents an unfortunate oversight in our experimental design. We located areas with structural correspondence, excised them, and then conducted sequencing. Following this, we compared the resulting VAF values to both the raw BaSISS signal VAFs and the model predictions formulated for the corresponding regions (refer to Figure 2.6).

The contrast between the noisy raw BaSISS signal VAFs and the nearly exact matching model predictions with the LCM-WGS estimates is striking. This contrast attests to the model's capability to precisely infer the spatial clonal composition of the tissue. Furthermore, it affirms the model's suitability for downstream data analysis.

2.4. Model assumptions and outlook

Quantitatively estimating the arrangement of genetically different clones in space, using information from their genes obtained through ISS, is a difficult task. The model described here addresses the multiple sources of biological and technical noise to generate clonal maps, which have been orthogonally validated (see Section 2.3). However, the design of any model inevitably introduces assumptions that must be considered during result interpretation. This section discusses the assumptions and limitations of the BaSISS model as well as possible future directions.

Signal binning

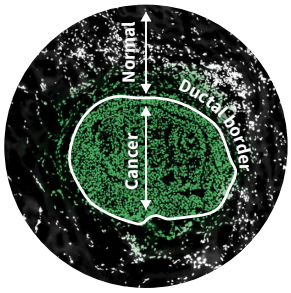
The BaSISS protocol relies on dividing ISS signals into regions, a process known as binning. This binning is convenient, as it reduces the number of parameters, enabling a straightforward application of the Negative Binomial likelihood function. The decision to bin is driven by the fact that the BaSISS signals are sparse; however, this leads to information loss. Although this trade-off suits the current stage of BaSISS experimental technology, it may become less justified as the technology evolves and improves.

Mathematical frameworks exist that allow the model to engage directly with the point signals. For example, both Markov Random Fields [Petukhov et al. 2022] and spatial Poisson processes [Qian et al. 2020] have been successfully employed to connect spots to the latent fields. Alternatively, the Integrated Nested Laplacian Approximation (INLA) [Rue et al. 2009] of the latent GP, although not yet popular in spatial genomics, appears promising.

I foresee that, as experimental technology advances, the BaSISS model can be extended to incorporate these approaches in the future.

Spatial correlation

The model employs a GP prior to handle the sparsity of the ISS data, assuming a smooth change in clonal composition across space. This assumption stems from the observation that breast cancer tissue maintains structural integrity, causing cells to remain in proximity after division. However, the model fails to consider detached compartments like ducts, leading to potential errors at the interfaces between compartmentalised regions. Consequently, the inferred clonal composition at these borders might appear more intermixed than in reality.



Ductal structure of the mammary gland provides a rigid border that prevents cancer cells from spilling out, yet the model incorrectly propagates information from the neighbouring green clone, leading to incorrect assumptions about the genetic status of the surrounding normal cells.

Copy number variation and expression

The BaSISS protocol detects RNA, while the model operates with clone genomes. To associate genetic clones with BaSISS data, I assume a positive linear relationship between an allelic variant's copy number and its expression. The correlation of BaSISS probes, normalised by WGS across the examined sections, endorses this assumption (see Section 2.3). Previous studies also support this relationship [Handsaker et al. 2015], though exceptions do exist.

For instance, genes such as *TP53* and *MYC*, as demonstrated by Shao et al. [2019], appear to deviate from the linear trend. This deviation may arise from dosage compensation mechanisms that can assume a non-linear form. An example of this can be seen in *MYC* expression, which seems to be resistant to low CNAs but exhibits high expression in focal high-level amplifications [Schukken et al. 2022]. These observations urge careful consideration during the design of allele target panels.

Homogeneity of clonal expression

The model assumes that alleles of a particular clone are expressed homogeneously across the slide. Potential violations might be partly compensated by multiple defining alleles of clonal branches, or by considering an inhomogeneous background (Section 2.2.4). The former might counterbalance the effects of highly variable alleles on the ELBO loss via overdispersion parameter, while the latter might capture some heterogeneity as pseudo-clones. Hence, the spatial distribution of the pseudo-clone factor may highlight problematic regions.

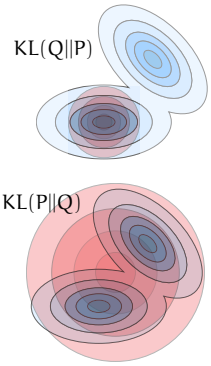
Phylogenetic tree

Since the clonal genotype matrix or phylogeny is fixed, the model cannot adjust it directly. Though selecting an optimal tree solution based on BaSISS data is theoretically feasible, in practice, various tree solutions yield similar ELBO values. This likely results from latent parameters causing non-identifiability. Attention should, therefore, be given to WGS mutation clustering and copy number estimation to avoid errors that may produce unrealistic clonal abundances and growth patterns. However, assessing model residuals could highlight problematic alleles and indicate the feasibility of the genotype matrix.

optimal tree solution *The original idea was to write the tree sampler and treat differences between ELBOs of the fitted models as the model selection criteria. This practice is justified in Chérief-Abdellatif [2018].*

Confidence intervals of the inference

The mean-field variational approximation offers a cost-effective means of inference, yet it comes with inherent limitations. Its first limitation arises from the aim of optimising KL divergence $\mathrm{KL}(\mathrm{Q}\|\mathrm{P})$. If the true posterior P displays multimodality, the variational approximation Q fits only one mode, consequently overlooking other viable solutions. An alternative definition of the reverse KL divergence might strive to encompass all modes, but it often proves undesirable, as the approximate distribution may assign high density to regions where the true posterior is improbable. Additionally, the mean-field approximation typically underestimates the parameter variance [Kucukelbir et al. 2016]. Regrettably, utilising full-rank VI or MCMC, which provide more accurate variance estimations, remains computationally impractical for the BaSISS model. Nevertheless technological developments in the field of high performance numerical computations and the use of more powerful hardware resources could make these approaches feasible in the future.



A common proverb postulates: "In the case of a non-alternative selection between two opposing entities, both will manifest suboptimally". Here, both formulations of KL *divergence exhibit significant drawbacks.*

Spatial Data Integration: Biological Insights from Multiple Data Modalities

3

Declaration

This chapter is based on the supplementary methods and technical results from:

Artem Lomakin, Jessica Svedlund, Carina Strell, Milana Gataric, Artem Shmatko, Gleb Rukhovich, Jun Sung Park, Young Seok Ju, Stefan Dentro, Vitalii Kleshchevnikov, Vasyl Vaskivskyi, Tong Li, Omer Ali Bayraktar, Sarah Pinder, Andrea L Richardson, Sandro Santagata, Peter J Campbell, Hege Russnes, Moritz Gerstung, Mats Nilsson and Lucy R Yates [Nov. 2022]. 'Spatial genomics maps the structure, nature and evolution of cancer clones'. en. In: *Nature* 611.7936, pp. 594–602.

The work I present here is primarily my own contribution. I developed and implemented the core mathematical and algorithmic toolkit to interpret BaSISS, ISS, IHC and histopathologic data under the supervision of M.Ger. and with inputs from A.S. and V.K..

I analysed and interpreted the data, and drafted the original article and figures, again under the supervision of M.Ger. and L.R.Y.. You can find all the code on Github. The code for hierarchical logistic regression I wrote together with V.K.. It is available at scHierarchy repo.

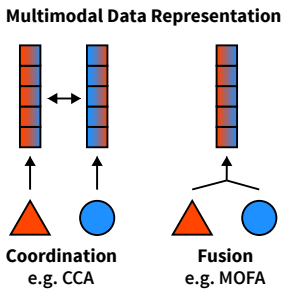
Another first author of this paper, J.S., in collaboration with P.J.C., M.N., and L.R.Y., designed the initial study of BaSISS and ISS. J.S., M.N., and C.S. conducted BaSISS, ISS and IHC experiments and provided the raw data. J.S.P, V.V., T.L., and M.Gat. preprocessed and decoded ISS and BaSISS data. I had access to the decoded ISS and BaSISS data. Hege Russnes provided histopathological description of the microregions. Sandro Santagata, Andrea L. Richardson, Sarah Pinder contributed histopathological expertise.

The Introductory Section 3.1, Data Characterisation Section 3.2.3, Hierarchical Logistic Regression Section 3.2.4, Multimodal Visualisation Section 3.2.8 and Summary Section 3.3 are original. I borrowed the rest of the sections from the original manuscript and reworked them to fit the thesis format, rewording, expanding and illustrating them for clarity. All the main and margin figures are original except Figure 3.1 and Figure 3.3, which I reproduced from the original paper with stylistic and compositional adjustments.

3.1. Background on multimodal analysis

Cancer progression involves complex interplay between the evolution of genome, cell state plasticity, spatial architecture, and interaction with the microenvironment (Chapter 1 and Section 4.1); the study of tumourogenesis therefore benefits from integrating multiple layers of genetic and non-genetic information at the single-cell level. Unimodal analyses, which focus solely on one aspect such as genetics or transcription, offer an incomplete view. To capture different axes of variation, integrative multi-omics approaches become essential [Baysoy et al. 2023; Nam et al. 2021].

Although ideal measurements would capture all modalities in the same cell simultaneously, current technologies fall short in their ability to deliver a truly multimodal view. Existing multi-omics technologies face limitations in throughput and the number of modalities they can measure concurrently [Baysoy et al. 2023]. Consequently, researchers often acquire unmatched single-cell information modalities across several biological samples. They then infer a joint representation in latent space, using techniques conceptually akin to canonical correlation analysis [Stuart et al. 2019] or factor analysis [Argelaguet et al. 2018; Velten et al. 2022]. However, this strategy suffers from inevitable information loss when aligning distinct feature spaces.



Two common ways of multimodal data representation that reflect cross-modal interactions. From Liang et al. [2022].

Current multi-omics methods typically measure two modalities, one of which is usually the transcriptome. Examples include CITE-seq (protein and transcription, Stoeckius et al. [2017]), scATAC-seq (DNA-accessibility and transcriptome, Lareau et al. [2019]), and G&T-seq (genome and transcriptome, Macaulay et al. [2015]). While these pairs are natural parts of the corresponding assays, they often exhibit low multiplexity and biases; for instance, CITE-seq generally measures only around 300 surface proteins. Nonetheless, one can use common dimensions as pivots for data integration and imputation, thereby better preserving information than when integrating over distinct feature spaces [Ashuach et al.

2023; Hao et al. 2021].

The methods outlined above offer a multimodal depiction of cells. However, because these assays are dissociative, they compromise the spatial structure. In contrast, spatial omics methods uniquely maintain the tissue context of cellular measurements (Box 1). Yet, they encounter similar challenges to those faced by single-cell multi-omics methods. Specifically, these methods share issues of sparsity and low plexity, and often necessitate integration with single-cell reference data obtained from dissociative studies (Box 2).

For spatial omics methods, integration of multiple spatially-resolved modalities is feasible across consecutive tissue sections (Box 2). Unlike in single-cell omics, where integration occurs for the same cell, spatial omics focuses on local neighbourhoods. This approach relies on the assumption that the cellular composition and properties remain relatively constant between adjacent slides. However, since most tissue are heterogeneous and contain functional structures, this assumption holds true only if the tissue sections are closely situated, typically within only a few cell layers. Algorithms for image registration enable the alignment of moderately distorted sections and the 3D reconstruction of a tissue block [Kiemen et al. 2020] unifying information obtained in different experiments. However, when integrating data, one must exercise caution to ensure that cellular structures have not vanished along the z-axis.

Spatial omics tools offer valuable context for cellular functions, exceeding the phenotypic descriptions available through traditional genomics. These tools elucidate the spatial organisation of cells within tissues, and its impact on cellular behaviour (Section 1.3). Analysis tools range from simple statistical tests that examine differences in cellular composition and expression to more complex unsupervised clustering algorithms based on topic models [Nirmal et al. 2022], Gaussian processes [Svensson et al. 2018] and spatial graphs [Danenberg et al. 2022]. For information-rich data, more explicit models of cell interaction, that attempt to model cell expression based on the local neighbourhoods, are possible [Fischer et al. 2023].

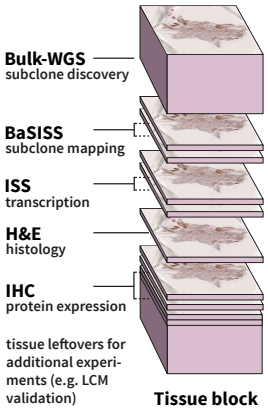
In summary, a nuanced understanding of cancer necessitates the integration of multiple data modalities, including spatial information. Although multi-purpose tools like Seurat [Hao et al. 2021], MOFA [Argelaguet et al. 2018], and MEFISTO [Velten et al. 2022] offer valuable platforms for data exploration, they may lack the specificity required to answer targeted questions in cancer biology, particularly when data quality is suboptimal. In this chapter, I will outline the spatial data collected alongside BaSISS, describe the integration methods employed to define and locate cell types, and present customised statistical tests designed to answer

the specific question: **"What are the phenotypic and microenvironmental differences between cancer clones?"**

3.2. Results

Knowing the spatial distribution of clonal densities as the result of BaSISS model described in Chapter 2, one may characterise the clones phenotypically based on additional spatially matched data. In this section, I outline the key properties of the data acquired to characterise the TME. I also detail the computational processing required to transform raw data into usable features. These include:

- Histopathological phenotype, Section 3.2.1
- IHC staining Section 3.2.2
- ISS expression signals assigned to nuclei, Section 3.2.3
- Cell type annotations, Sections 3.2.4 and 3.2.5



The tissue block undergoes sectioning to yield material for the designated experiment. Subsequently, the various data modalities are spatially aligned. The thickness of each slide measures 10 µm.

The modalities of spatially resolved information exist across multiple consecutive tissue slides. These slides exhibit spatial deformations from one layer to another, complicating the registration process. To integrate information across all modalities, a trained pathologist, Hege Russnes, manually annotated regions of breast tissue. These annotations targeted areas with similar histopathological phenotypes, which should be in close physical proximity on the z-stack. After integrating the data, I applied statistical tests specifically designed for this dataset to characterise cancer biology (Section 3.2.6).

3.2.1. Histopathological phenotype

In our study, histology served dual purposes: it helped define micro-anatomical and cellular features, and designate microregions for feature projection across consecutive slides and assays (Box 1). Hege Russnes annotated microregions on haematoxylin and eosin (H&E) stained fresh-frozen samples, uninfluenced by prior knowledge of the sample's genetic clonal structure.

For ductal carcinoma *in situ* (DCIS) cases, annotation focused on regions presumed to be myoepithelially enclosed, as indicated by H&E or other histological images (refer to Figure 4.1 for details). Due to the limitations imposed by fresh freezing, certain features like mitotic activity and chromatin structure were difficult to assess. Consequently, nuclear grading (see Section 4.1) in DCIS relied

mainly on nuclear size. H.R. assessed the presence or absence of nuclear vacuoles using a 20% cut-off criterion. Challenges in identifying terminal ductal lobular unit (TDLU) and ducts were resolved by considering the presence of branches in morphology (Figure 4.1).

For invasive tumour sections lacking clear histological structures, H.R. and L.R.Y. manually demarcated similarly sized regions of invasive tissue. This approach enabled uniform sampling across the entire tissue surface area.

In specific samples, P2-TN1 and P2-LN1, H.R. reported multiple independent growth patterns. For P2-LN1, areas were selected based on these growth patterns, with the aim of creating regions of similar size (Figure 3.1). In the case of P2-TN1, where different growth structures were less distinctly demarcated, H.R. reported the proportion of cells belonging to each growth pattern within predefined regions.

3.2.2. Immunohistochemistry data

Immunohistochemistry serves as a robust method for localising specific proteins within tissue sections. While the technique offers limited multiplexing capabilities, it remains a standard approach for obtaining spatial information on protein expression levels.

Carina Strell incubated tissue sections with antibodies targeting PanCK, CD45, and HER2. These slides underwent both BaSISS and ISS protocols. In subsequent slides, C.S. and S.S. targeted additional proteins, such as SM-MHC, P63, PR, Ki-67, and PTEN. For further context on the role of these proteins in breast cancer, refer to Section 4.1.

L.R.Y. utilised Qupath digital software for the quantitative evaluation of stained nuclei [Bankhead et al. 2017]. The analysis focused on assessing the expression of Ki-67, PTEN, and PR in selected regions (Appendix Figure D.1).

3.2.3. *In situ* sequencing data

The study design employs In Situ Sequencing (ISS) technology to characterise the transcriptional phenotype of tissue samples. The ISS dataset included two panels: a previously published 91-gene panel targeting oncology-related markers and a novel 62-gene immune panel designed for immune cell profiling (Figure 3.1). I assessed the immune panel's performance by correlating its expression levels with paired bulk RNA-seq data in the triple-negative breast cancer (TNBC) case outlined in Section 4.2.2. The choice of this specific case stemmed from its el-

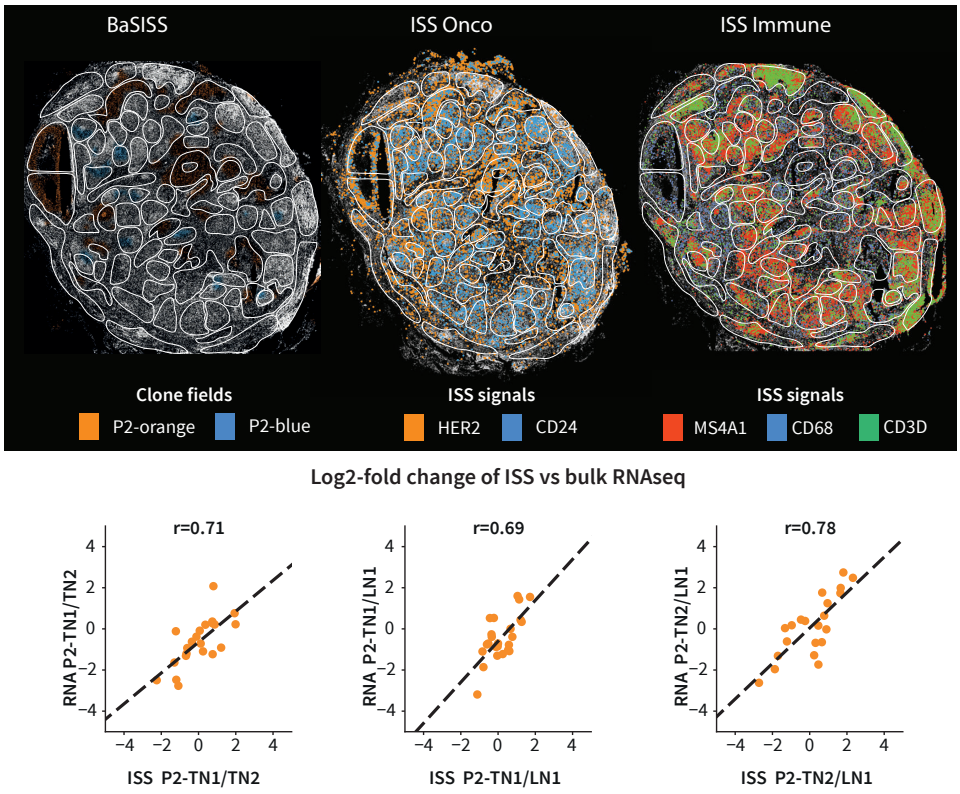


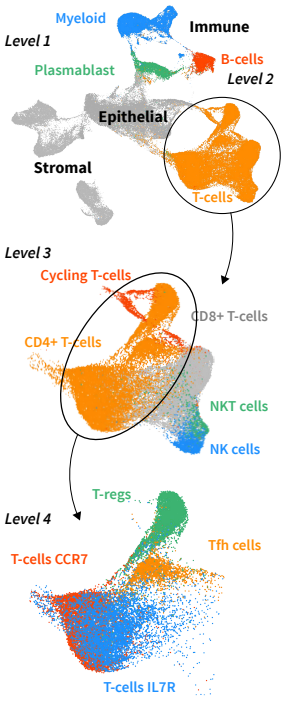
Figure 3.1: *in situ* sequencing data for the lymph node breast cancer sample. ***top***, Display of BaSISS fields alongside two selected ISS panels, focusing on selected genes (*HER2*, *CD24* - oncology panel and *MS4A1*, *CD68*, *CD3D* - immune panel). White outlines mark microregions, identified by a histopathologist, based on histological growth patterns. ***bottom***, Scatterplots show the $\log_2$-fold change in gene expression between samples, as measured by RNAseq and combined ISS oncology and immune experiments. High correlation values (R = Pearson's correlation) confirm on-target probe performance. The analysis includes genes that meet the following criteria: a Transcripts Per Million (TPM) count greater than 25 in RNAseq and over 1000 detections per million cells in ISS, minimising deviations due to low counts.

evated levels of immune cells, the panel's designed target. Once normalised for technological differences, data exhibited a significant correlation (R = 0.69–0.78, Pearson's) as shown in (Figure 3.1).

Transcript detection, particularly differential expression near clones, often provides functional insights if a gene has a clear association with a specific cell type. For instance, the presence of the *MS4A1* gene strongly indicates the existence of B-cells. However, many genes lack a clear association with a cell type, complicating interpretation. The detection of *ACTA2*, for example, could signify the presence of either smooth muscle cells, myofibroblasts, or basal epithelial cells. Additionally, functional interpretation is confounded by the cell signal's origin; epithelial-mesenchymal transition (EMT) markers in epithelial cells signify transition, whereas the same markers in stromal cells are standard. Therefore, accurate interpretation of ISS data requires knowledge of the originating cell type.

While the ISS panel aimed to identify specific cell types and was partially based on the Oncotype DX panel [Paik et al. 2004], the gene selection process did not consult single-cell atlases. Instead, the chosen genes were primarily informed by the existing literature, which often focuses on protein characteristics unmeasurable by ISS. Moreover, the data was generated using an older version of the ISS technology based on sequencing by ligation, resulting in a low signal count per cell (fewer than 3 signals per cell when assigned) (Appendix A.5)

Given the data's sparsity, the use of sophisticated methods designed to resolve niche cell types, such as cell2location [Kleshchevnikov et al. 2022] and SSAM [Park et al. 2021], based on expression signatures turned out to give suboptimal results. Consequently, I opted for an alternative approach: employing scRNA-seq data to first identify cell markers that unequivocally define cell types, and subsequently using these markers for tissue profiling.



Cell-type organisation inherently exhibits a hierarchical structure, often reflected in single-cell atlases annotations. Data from S. Z. Wu et al. [2021]

3.2.4. Hierarchical logistic regression for cell marker discovery

The objective is to identify easily usable cell type markers for classifying individual cells within a tissue. To make decoding simple, these markers should be unique to a specific cell type or, alternatively, a unique combination of markers should denote a specific cell type. Identifying unique markers is challenging because most genes are involved in cellular programmes that multiple cell types share, particularly when the classification is granular. Current methods for discovering cell type markers rely on statistical tests [F. A. Wolf et al. 2018], heuristic approaches for dimensionality reduction [Dai et al. 2022; Dumitrascu et al. 2021;

Missarova et al. 2021], or machine learning classification models [Nelson et al. 2022] coupled with feature importance analysis. These methods generally fail to explicitly incorporate the concept of combinatorial encoding of cell types.

One natural approach to improving cell type identification exploits the hierarchical organisation of cell types. In this model, we assume that higher level group cell markers, such as those for all epithelial or immune cells, are obligatorily expressed across their respective subtypes. Nonetheless, we permit subtypes to display non-unique markers that might be shared among multiple groups as soon as this group belongs to a different overarching type. The key requirement is that this higher level cell type remains identifiable.

Consider, for example, the immune-cycling and epithelial-cycling cell states. Both types of cells express markers indicative of cell cycling. Non-hierarchical marker discovery methods may overlook these cycling markers as they are too general. This limitation arises because subpopulations within both immune and epithelial cells express these markers. However, by combining cycling markers with high level markers specific to all immune cells, one can unambiguously categorise immune-cycling cells.

To formalise this strategy, one could employ a marker selection model like logistic regression restricting the model to groups of cells that belong to the same high level cell type. Then, iterate this approach across multiple layers of hierarchy. However, this isolated approach hinders information sharing across hierarchical levels and could yield suboptimal results. To address this limitation, I have restructured the standard logistic regression model to operate within the cell hierarchy framework, aiming to identify cell type markers at each hierarchical level (Figure 3.2).

Multi-class logistic regression

Consider a Bayesian formulation of the logistic regression model for predicting cell types. Let y_i represent the cell type label for cell i, which can assume one of K possible classes, $\{1, 2, \ldots, K\}$. Define $\mathbf{x}_i$ as the g-dimensional vector of gene expression levels for cell i. The likelihood function for this model is given by:

$$p(y_i = k|\mathbf{x}_i, \mathbf{w}) = \frac{\exp(\mathbf{w}_k^T \mathbf{x}_i)}{\sum_{k=1}^{K} \exp(\mathbf{w}_k^T \mathbf{x}_i)} \tag{3.1}$$

Here, $\mathbf{w}_k$ is the weight vector corresponding to class k. To promote sparsity, we employ a zero-centred Laplace prior for the weights, characterised by scale parameter λ:

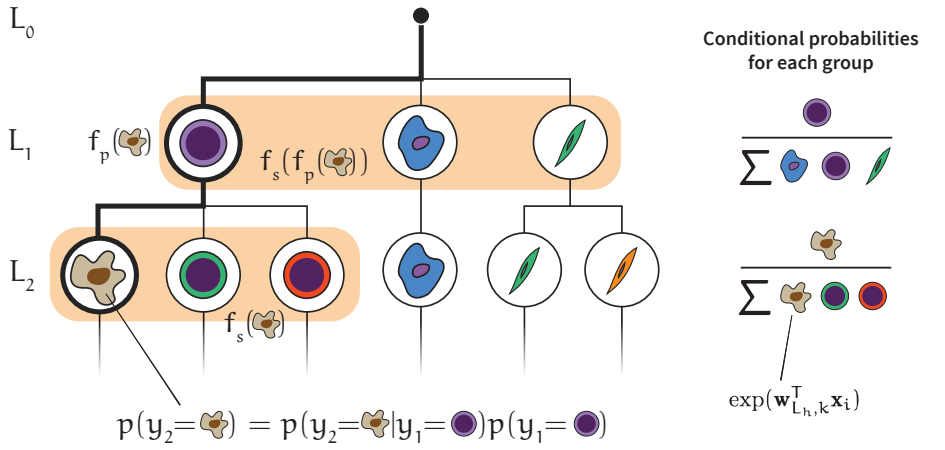


Figure 3.2: Hierarchical logistic regression conceptual description. The hierarchical logistic regression model utilises a tree that represents the hierarchical relationships among different cell types. Each node in this tree corresponds to a specific cell type, and edges signify subclass relationships. Mathematically, the likelihood conditions the probability of a cell belonging to a specific type based on its hierarchical ancestors, leveraging an array of weight vectors corresponding to each hierarchical layer. This structure introduces conditional dependencies between cell types, allowing contextual cell-type marker predictions

$$p(w_{kg}) \sim \text{Laplace}(0, \lambda), \quad \text{i.i.d. for all } g \tag{3.2}$$

I fit this model to an annotated scRNA-seq dataset using the VI method, as detailed in Section 2.1.3. This provides an approximation of the posterior distribution $p(\mathbf{w}_k|\mathbf{X}, \mathbf{y})$. The magnitude of each $w_{k,g}$ thereby affords a straightforward interpretation of gene g's importance in classifying cell type k.

Graph structure of annotated single-cell data

The graph structure of single-cell data annotation hierarchies can be mathematically represented as a tree $T = (V, E)$, where V is the set of vertices and E is the set of edges. Each vertex $v \in V$ represents a particular cell type annotation, and each edge $(u, v) \in E$ represents a subclass relationship between cell types u and v.

Let r be the root node, representing the most general cell type encompassing all cells. Nodes directly connected to r represent broad cell types; for example these could be "Epithelial cells", "Immune cells" or "Stromal cells". The set of all nodes connected to the root is denoted as L_1 and represents the first layer of the

hierarchy.

Mathematically, a hierarchical layer L_i consists of nodes that are equidistant from the root node r. That is, for any node v in L_i, the shortest path from r to v has length i:

$$L_i = \{v \in V : d(r, v) = i\} \tag{3.3}$$

where $d(r, v)$ denotes the shortest path from r to v in tree T.

For convenience, let's also define a function f_p (Figure 3.2) that retrieves the parent u of the child node v:

$$f_p(v) = u, \text{ where } (u, v) \in E \tag{3.4}$$

The f_s function (Figure 3.2) returns the set of siblings of node v (including v itself), that lie in the same layer as v node, and belong to the same parent node u:

$$f_s(v) = \{w \in V : f_p(w) = f_p(v)\} \tag{3.5}$$

Taking an element in L_3, and iteratively applying f_p until we reach L_1 would return it's hierarchy "CD8+ T cells" $\rightarrow$ "T cells" $\rightarrow$ "Immune cells". Similarly, applying f_s to this element would return all other cell types in that belong to the same parent node "T cells", such as "$CD4_+$ T cells" and "NK cells".

Hierarchical extension of logistic regression

Now that we have defined the hierarchical structure of the cell type annotation, we can extend the logistic regression model to incorporate this information (Figure 3.2). Similar to standard logistic regression, $\mathbf{x}_i$ as the g-dimensional vector of gene expression levels for cell i. Let $\mathbf{y}_i = [y_{i,1}, y_{i,2}, \ldots, y_{i,H}]$ be the set of labels for cell i, corresponding to each of the H hierarchical layers. Similarly, the weight vector for layer L_h is denoted as $\mathbf{w}_{L_h}$, where $h \in 1, 2, \ldots, H$.

The probability of cell i belonging to class k at layer L_h is now dependent on the hierarchical parents of this cell:

$$\begin{aligned} p(y_{i,h} = k) = p(y_{i,h} = k | y_{i,h-1} = f_p(k)) \\ \times p(y_{i,h-1} = k | y_{i,h-2} = f_p(f_p(k))) \\ \times \cdots \\ \times p(y_{i,h_1} = f_p^{H-1}(k)) \end{aligned} \tag{3.6}$$

Notice that the conditional term like $p(y_{i,h} = k | y_{i,h-1} = f_p(k))$ is equivalent

to the standard logistic regression model, restricted to the $f_s(k)$, i.e. the siblings of label k.

Expanding the above equation using Equation (3.1) yields the likelihood function for the hierarchical logistic regression model:

$$\begin{aligned} p(y_{i,h} = k|\mathbf{x}_i\mathbf{w}) = & \frac{\exp(\mathbf{w}_{L_h,k}^T \mathbf{x}_i)}{\sum_{j \in f_s(k)} \exp(\mathbf{w}_{L_h,j}^T \mathbf{x}_i)} \\ & \times \frac{\exp(\mathbf{w}_{L_{h-1},f_p(k)}^T \mathbf{x}_i)}{\sum_{j \in f_s(f_p(k))} \exp(\mathbf{w}_{L_{h-1},j}^T \mathbf{x}_i)} \\ & \times \cdots \\ & \times \frac{\exp(\mathbf{w}_{L_1,f_p^{H-1}(k)}^T \mathbf{x}_i)}{\sum_{j \in L_1} \exp(\mathbf{w}_{L_1,j}^T \mathbf{x}_i)} \end{aligned} \tag{3.7}$$

The prior distribution for the weights $\mathbf{w}$ is the same Laplace distribution as in Equation (3.2). To address the issue of class imbalance, we now adjust the weights using the square root of the number of cells in each group. We employ square root scaling to moderate the penalty on larger groups, thereby maintaining a balanced emphasis across different classes.

This model is fit using ADVI, and the posterior distributions of the weights are used to asses the importance of a gene to define a given class.

Inferred set of markers for ISS panels

I run the aforementioned model on the breast cancer atlas derived from scRNA-seq data (~ 30000 genes), as described by S. Z. Wu et al. [2021]. The analysis was limited to the subset of genes present in two ISS panels (91 + 62). After applying a filter to retain only the top 70^{th} percentile of all gene weights, and manually excluding genes expressed in multiple cell types within each group, the resulting marker set exhibited limited resolution (Figure 3.3). Specifically, the markers were able to differentiate only two out of the four hierarchical levels defined in the atlas. At the first level, adequate markers allowed for differentiation among broad cell types: "Immune", "Epithelial", and "Stromal". At the second level, the "Immune" cells resolved further into "B-cells", "Myeloid", and "T-cells". The "Stromal" category subdivided into a mixed population of "Fibroblasts + PVL" and "Endothelial" cells. The analysis did not provide further resolution due to a poor choice of markers during panel design.

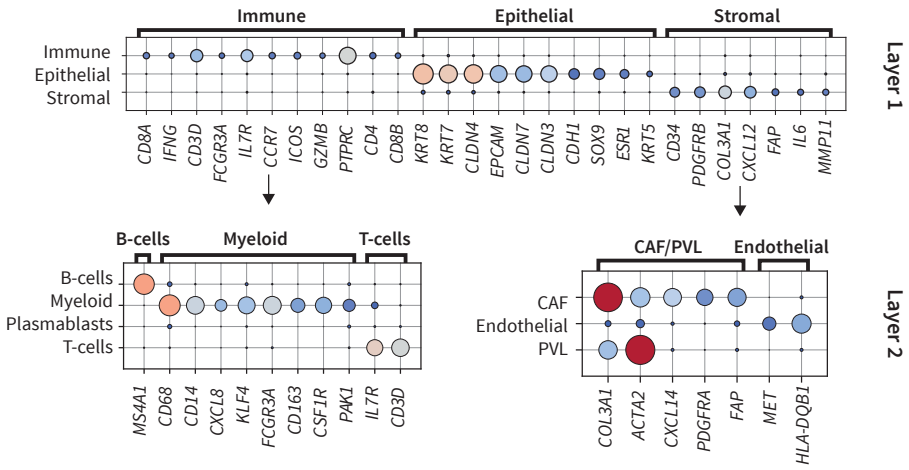


Figure 3.3: Hierarchical cell-type markers for ISS panels. The result of running hierarchical logistic regression to select marker genes for cell typing, employing Breast Cancer scRNA-seq Atlas [S. Z. Wu et al. 2021] as input dataset. The analysis was limited to genes from ISS oncology and immune panels. The figure displays genes with the highest weights for each hierarchical layer. PVL = perivascular-like cells. CAF = cancer associated fibroblasts

3.2.5. Cell type assignment with sparse data

I overlaid the ISS data with nuclei segmentation masks and allocated signals to the closest nuclei. I applied a conservative distance cut-off of 5µm — approximately double the nuclear radii — to minimise the likelihood of misannotation, resulting in approximately 30% signal loss.

Cell-type classification occurred in two steps in a simple 'if-else' algorithm. In the first iteration, nuclei possessing any markers corresponding to the broad categories ("Epithelial", "Immune", "Stromal") were assigned accordingly. In the subsequent iteration, nuclei initially classified as "Immune" or "Stromal" underwent further categorisation based on the presence of specific markers shown on Figure 3.3. Nuclei lacking proximal markers or exhibiting conflicting assignments received an "Unassigned" classification.

Generally, only approximately 2–40% of nuclei were categorised into even the highest level cell types, with notable variations across samples (Table 3.1). Such severe technical variability precludes the possibility of making biologically relevant comparisons between samples. Consequently, the analysis discussed in Section 3.2.6 and Chapter 4 focus solely on intra-sample comparisons. The annotated maps for the lymph node case are presented in Figure 4.8 and Appendix

Figure D.5.

Table 3.1: Proportions of nuclei classified by cell type across samples. This table presents the results of cell-type classification, achieved through the use of unambiguous marker genes and a straightforward 'if-else' algorithm. The table focuses on higher level categories, although "Immune" and "Stromal" cells undergo further subclustering. The generally low proportions of identified cell types are attributable to the sparse nature of ISS data, suboptimal panel design and cautious signal attribution to nuclei.

Sample ID	ISS panel	Immune	Epithelial	Stromal	Unassigned	Total Cells
P1-D1	Onco	0%	1.4%	0.3%	98.4%	519,438
	Immune	1.7%	0%	0.7%	97.6%	622,534
P1-ER1	Onco	0%	7.7%	11.7%	80.6%	286,153
	Immune	6.5%	0%	4.4%	89.1%	275,000
P1-ER2	Onco	0%	3.8%	4.2%	92.0%	322,733
	Immune	4.7%	0%	2.3%	93.0%	318,678
P1-D2	Onco	0%	19.2%	4.1%	76.7%	195,726
	Immune	4.8%	0%	4.4%	90.8%	159,537
P1-D3	Onco	0%	9.8%	5.2%	85.0%	104,651
	Immune	7.8%	0%	3.7%	88.6%	122,455
P2-TN1	Onco	0%	8.6%	22.8%	68.5%	243,013
	Immune	9.7%	0%	1.6%	88.7%	231,996
P2-TN2	Onco	0%	20.7%	6.0%	73.3%	267,044
	Immune	6.1%	0%	1.2%	92.8%	279,053
P2-LN	Onco	0%	5.0%	22.6%	72.5%	400,405
	Immune	15.4%	0%	1.6%	83.0%	410,762

3.2.6. GLMM for multiregional quantitative analysis of clone-specific differences

Upon allocating ISS signals to individual cells and identifying cell-types, microregions contain several types of information. These include continuous clonal contributions, counts of IHC-stained nuclei, ISS signals categorised by originating cell type, counts of each cell type, and categorical histological phenotype features (Section 3.2).

The primary objective is to elucidate the phenotypic, transcriptional, and compositional differences between clones. It appeared that breast cancer clones generally do not intermingle, as discussed in Sections 2.3 and 4.2.5 and shown on Figures 4.7 and 4.9. Thus, I restricted the analysis to highly clonal regions based on specific CCF thresholds: $CCF_{clone} > 0.7$ for P1, and $CCF_{clone} > 0.15$ and 0.05 for P2-TN1 and P2-LN1, respectively. The lower thresholds for P2-TN1 and P2-LN1 account for the high levels of non-epithelial cells present in these samples.

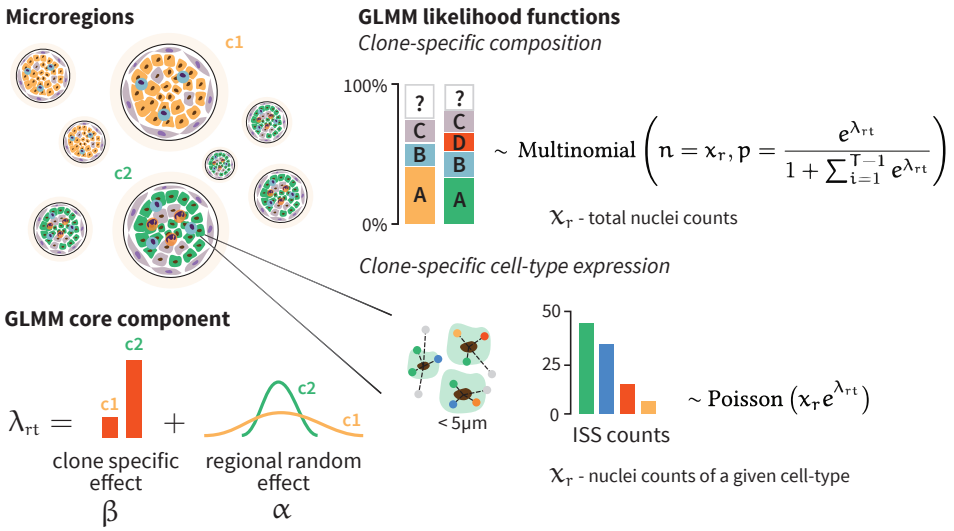


Figure 3.4: Modelling clone-specificepxression and composition using GLMM. On of the objectives of this study is to characterise the phenotypic, transcriptional, and compositional differences among cancer clones. I record cell-type composition and ISS signal counts, attributing them to cells with known cell types. I then model these data using a generalised linear mixed model (GLMM) that incorporates a clone-specific fixed effect, denoted as β, and a clone-specific regional random effect, denoted as α. Inferred posterior distribution of β is then used to compute statistics and reason about differences between clones.

Standard statistical tests like Fisher's exact test easily handle categorical data, such as histological annotations. However, these methods are less applicable for count data, which are subject to variations in cell numbers and likely overdispersion (e.g. region-specific expression variation). To address these challenges, I employed custom generalised linear mixed models (GLMMs) designed to account for variable cell numbers and data overdispersion (Figure 3.4).

Intrinsic expression in cell types

To quantify differential expression associated with clonal regions for a specific cell type, we record ISS signals corresponding to nuclei classified under that cell type. I model the distribution of observed ISS expression signals Y_{rg} for genes g in regions r using a clone-specific expression rate β_{cg}. This rate is sampled uniformly, encompassing all plausible values (after exponentiation):

$$\beta_{cg} \sim \text{Uniform}(-10, 10) \tag{3.8}$$

Acknowledging the possibility of region-specific variation in expression, let's introduce a random effect variable α_{rg}. This variable has a clone-specific hierarchical prior σ_{cg}, which governs the prior belief regarding the extent of inter-regional variability. A value of 0.05 for this hyperprior anticipates low variability, yet permits deviations if the data suggest otherwise. The binary matrix A_{rc} maps dominant clones to their corresponding regions.

$$\sigma_{cg} \sim \text{HalfNormal}(0.05) \tag{3.9}$$
$$\alpha_{rg} \sim \text{Normal}(0, A_{rc}\sigma_{cg}) \tag{3.10}$$

After adjusting for the number of nuclei of the relevant cell type x_r, the likelihood becomes:

$$Y_{rg} \sim \text{Poisson}\left(x_r e^{A_{rc}\beta_{cg}+\alpha_{rg}}\right) \tag{3.11}$$

This formulation accommodates both clone- and region-specific variations in gene expression, providing an interpretable statistical model for analysing cell-type specific differential expression (Figure 3.4).

Cell-type composition

To assess differential cellular composition between regions associated with clones, I computed the counts of nuclei categorised as specific cell types. I then modelled the distribution of nuclei counts Y_{rt} attributed to cell type t across regions r as a mixed model. This model incorporates both a clone-specific frequency β_{ct} and a region-specific random effect α_{rt}, akin to the approach used for **cell-type specific expression**:

$$\beta_{ct} \sim \text{Uniform}(-10, 10) \tag{3.12}$$
$$\sigma_{ct} \sim \text{HalfNormal}(0.05) \tag{3.13}$$
$$\alpha_{rt} \sim \text{Normal}(0, A\sigma_{ct}) \tag{3.14}$$

Given the sum-to-one constraint on cell type frequencies, I employed a softmax (logit) link function along with a Multinomial likelihood for the cell types:

$$\lambda_{rt} = A\beta_{ct} + \alpha_{rt} \tag{3.15}$$

$$Y_{rt} \sim \text{Multinomial}\left(n = x_r, p = \frac{e^{\lambda_{rt}}}{1 + \sum_{i=1}^{T-1} e^{\lambda_{rt}}}\right) \tag{3.16}$$

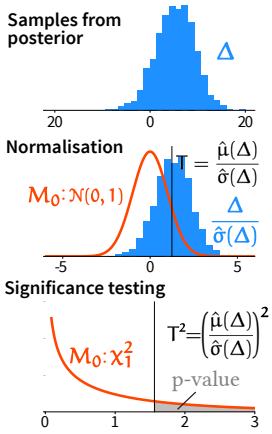
By incorporating these constrains, I ensure that the model accounts for the inherent constraints and random effects in differential cellular composition across regions (Figure 3.4).

IHC staining

L.R.Y. counted the number of IHC-positive and IHC-negative nuclei in each region using QuPath [Bankhead et al. 2017]. To enable the application of the aforementioned **cell-type composition** model, I treated IHC^+ and IHC^- nuclei as two distinct cell types.

Inference and statistical testing

I implemented the models using the NumPyro library and employed Hamiltonian Monte Carlo (HMC) for parameter inference (Section 2.1.3). To identify quantitative differences, I performed pairwise comparisons of clone-specific expression rates β_{cg}, for each pairs of clones. Examples of posteriors for expression rates appear in Appendix Figure D.1



We approximate the posterior of expression differences between clones with a χ^2_1 distribution for more precise quantile estimates than MCMC sampling alone provides

HMC generates empirical posterior distributions with a finite number of samples, limiting our ability to resolve small differences. Specifically, the quantiles are restricted to values of $1/n$, where n is the sample size from the HMC run (I run with n=4000). This limitation becomes particularly challenging when conducting multiple comparisons, as it necessitates the multiple testing correction.

To enable the calculations of more extreme quantiles I propose a hypothesis test based on the idea that the posterior of expression differences between clones, denoted as $\Delta_g = \beta_{1g} - \beta_{2g}$, should have a mean of zero under the null model M_0 (similar to H_0 in frequentist analysis). Assuming that Δ_g is normally distributed and treating the posterior as analogous to the frequentist test statistics, let's introduce T, essentially a z-score for the mean, defined as:

$$T = \frac{\hat{\mu}(\Delta)}{\hat{\sigma}(\Delta)} \tag{3.17}$$

Under M_0, T should be approximately normally distributed with a mean of zero and a standard deviation of one. Consequently, the square of the test statistic T^2 should follow a chi-squared distribution with one degree of freedom,

denoted as χ^2_1, under M_0. The use of T^2 allows for a test of magnitude rather than direction of the differences. To calculate the p-value, I use $1 - \text{CDF}(\chi^2_1)$. If necessary, I apply multiple testing corrections to each comparison to maintain an FDR below 0.1.

P-values computed according to this approach appear on the volcano plots on Figure 4.11 and Appendix Figures D.2 and D.4. For differential composition and protein expression, I used a similar approach with the softmax-transformed β_{ct} values. Examples of β_{ct} posteriors and differential composition statistics appear in Figure 4.11 and Appendix Figures D.1 and D.2.

To maximise statistical power in clone-specific differential expression analysis, I excluded genes with a clone-agnostic average number of detected signals per region less than 0.3.

3.2.7. Approach limitations

The algorithms discussed in this chapter aim to integrate multiple data layers to answer specific biological questions. While they mitigate issues like data scarcity and permit spatial integration of deformed tissues, they come with inherent limitations and assumptions that must be considered in result interpretation.

Firstly, the current signal assignment is inefficient. It is limited to a spatial resolution of 5 µm, leading to the loss of 30% of all signals. Optimally, a probabilistic method like pciSeq [Qian et al. 2020] or Bayesor [Petukhov et al. 2022] should be employed. Unfortunately, the limited number of signals precludes the use of these methods. This limitation may be overcome as the number of signals increases.

Secondly, I employed simple cell-type assignment algorithms out of necessity, rather than as best practice, and the failure of probabilistic algorithms led to this choice. While similar 'if-else' assignments exist in multiplex IHC-based technologies with gating, these usually rely on aggregate signals, not individual ones. Such decision-making based on a limited number of signals risks generating false positives.

The decision to manually define discrete microregions, rather than automatically identifying histological units, or working in a continuos space, arose from challenges such as poor cell-type identification and inconsistent tissue deformations between slides. These issues made the use of registration algorithms problematic, complicating the integration of data layers across spatial dimensions. On one hand, this manual approach simplified analysis by providing clearly defined histogenomic entities as a basis for statistical inference. On the other hand, it in-

troduced bias and led to the exclusion of regions omitted from analysis for reasons such as ambiguous spatial structure, presence of mixed clones, or even fatigue on the part of the histopathologist — a concern that is more significant than it may initially appear.

Lastly, the approach assumes uniformity in the marker gene expression within the same annotation group. Although the regional random effect variable α captures some regional variability, it does not account for systemic biases in cell group expression. For instance, if high level classes like immune cells express markers differently between subtypes (e.g. B-cell vs Myeloid), this will affect the frequency of their identification, thereby confounding the bespoke compositional analysis.

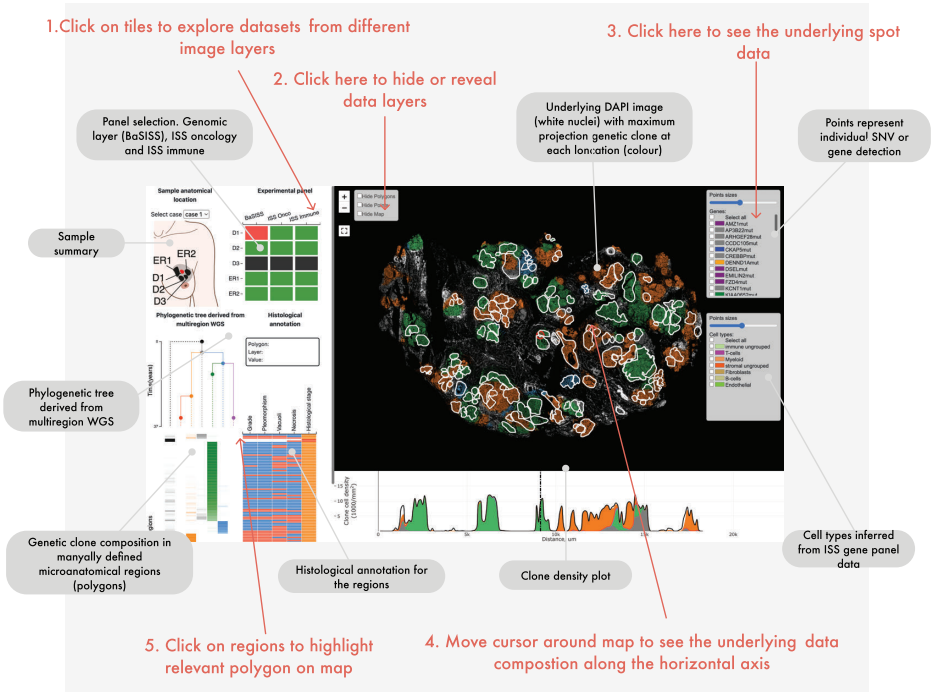


Figure 3.5: Multimodal data visualisation web-tool at cancerclonemaps.org. The web tool visualises multimodal data, specifically designed to meet the requirements of the BaSISS experiment. This web tool interactively displays microregions border, histological features, ISS and BaSISS data, inferred clonal maps, and ISS-based cell type locations for all samples analysed in the study from Lomakin et al. [2022]

3.2.8. Multimodal data visualisation

The challenge of visualising spatial multimodal data, particularly in the project that combines BaSISS, ISS, and histological data, requires a tailored approach. Current methods such as TissUUmaps [Pielawski et al. 2023], OMERO [Allan et al. 2012], WebAtlas [T. Li et al. 2023], and Napari [Ahlers et al. 2023] fall short in meeting specific visualisation requirements, especially for clone fields inferred from BaSISS data.

To address this challenge, I created a specialised web-based tool together with Gleb Rukhovich and Lucy R. Yates, available at cancerclonemaps.org. Our tool utilises a Flask backend and incorporates frontend technologies like D3 for visual elements, Pixi.js for efficient spot visualisation, Leaflet for tissue mapping, Geoman for map selection tools, and Plotly.js for the diagrams. This stack of technologies enables interactive exploration of our complex data sets, offering a unified platform to investigate histogenomic relationships (Figure 3.5). Consequently, this tool streamlines exploratory analysis and assists in identifying qualitative patterns across different data types.

3.3. Summary

In Chapter 2, I detail experimental methods and statistical algorithms for identifying genomically-defined cancer subclones. Chapter 3 extends this discussion by explaining how to integrate genetic information with spatial data on transcription, proteomics, and histology. By combining these approaches, the assembled toolkit aims to shed light on both the intrinsic cellular properties and the tumour microenvironment. Specifically, it seeks to identify differences between cancer clones in terms of their intrinsic features and their respective habitats.

While some assays discussed here lack multiplexing capability and may appear outdated in the fast-evolving field of spatial genomics, they can still provide valuable insights. When interpreted carefully, the generated data can reveal intricate details about the spatial evolution and ecology of tumours. The following section (Chapter 4) demonstrates the utility of these methods through a case study on breast cancer.

Spatial Evolution and Ecology of Breast Cancer: From Ducts to Lymph Nodes

4

Contributions

This chapter is largely based on the main analysis from:

Artem Lomakin, Jessica Svedlund, Carina Strell, Milana Gataric, Artem Shmatko, Gleb Rukhovich, Jun Sung Park, Young Seok Ju, Stefan Dentro, Vitalii Kleshchevnikov, Vasyl Vaskivskyi, Tong Li, Omer Ali Bayraktar, Sarah Pinder, Andrea L Richardson, Sandro Santagata, Peter J Campbell, Hege Russnes, Moritz Gerstung, Mats Nilsson and Lucy R Yates [Nov. 2022]. 'Spatial genomics maps the structure, nature and evolution of cancer clones'. en. In: *Nature* 611.7936, pp. 594–602.

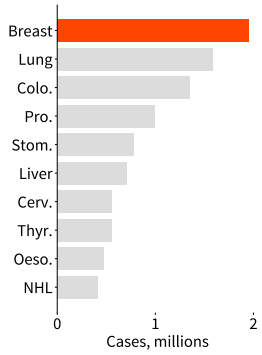
The work I present here is primarily my own contribution. I focused on developing and implementing the core mathematical model for the BaSISS data under the supervision of M.Ger., with valuable inputs from A.S. and V.K.. I analysed and interpreted the data, and drafted the original article and figures together with and under the supervision of M.Ger. and L.R.Y.. You can find all the code on Github.

Another first author of this paper, J.S., in collaboration with P.J.C., M.N., and L.R.Y., designed the initial study of BaSISS. J.S., M.N., and C.S. conducted the experiments and provided the raw BaSISS data. J.S.P, V.V., T.L., and M.Gat. preprocessed and decoded this data. L.R.Y. performed the LCM cuts, and S.D. conducted the WGS subclonality analyses. I had access to the decoded BaSISS and LCM-WGS data.

The Introductory Section 4.1 is original. I borrowed the rest of the sections from the original manuscript with only minor stylistic and structural alterations. All the main and margin figures in the introduction are original. Figures Figures 4.7

to 4.11 and the margin figure in Section 4.2.5 are borrowed from the original paper with minor stylistic and compositional adjustments.

4.1. A Primer on Breast Cancer



Breast cancer is the most prevalent type of cancer world wide, with approximately 2.2 million cases registered in 2020.

In 2020, breast cancer surpassed lung cancer to become the most diagnosed cancer type in the world [Sung et al. 2021]. It accounted for approximately 11.7% of all diagnosed cancer cases. This trend occurred irrespective of a country's development index, predominantly targeting the female population. Within this demographic, breast cancer constituted nearly a quarter of all diagnoses, and 15.5% of all cancer-related deaths. An estimated 2.2 million cases were registered in 2020, and this number is expected to grow, partly due to obesity, the ageing population and the fact that females constitute the majority of the older population. Despite advances in treatment and early detection, the death rate is not declining rapidly enough to counteract the consistent 0.5% increase in incidence rate observed over the past two decades. [Cronin et al. 2022]. This motivates the need for a better understanding of the disease's biology and evolution.

4.1.1. Normal breast tissue structure and development

placodes *thickened areas of ectodermal tissue in embryonic development that give rise to specialised structures.*

Before exploring the biology of breast cancer, it is crucial to understand the biology of the affected organ - the mammary gland. Development of the mammary gland begins in the embryo through interactions between epithelial and mesenchymal cells. At first, mammary placodes form and subsequently invaginate to create mammary buds. After that, the ectoderm proliferates and branches into the fat pad to form a rudimentary ductal tree. Throughout the entire process, paracrine signalling between the ectoderm and surrounding mesenchyme is crucial, as epithelial cells induce specialised mesenchymal cells formation, that in turn instruct the ectoderm to commit to mammary identity [Macias et al. 2012; Spina et al. 2021].

The mature ductal epithelium consists of two distinct cell types: luminal cells, which express KRT7, KRT8, and KRT18, and basal cells, marked by KRT5 and KRT14 expression. Luminal cells constitute the inner layer and play a critical role in milk production and duct formation. In contrast, basal cells form the outer layer and adhere to the basement membrane. They execute the contractile function essential for milk release. During embryogenesis, mammary epithelial cells undergo lineage restriction. Early bipotent progenitors express both basal and luminal markers [Prater et al. 2014; Rios et al. 2014; Stingl et al. 2001]. As devel-

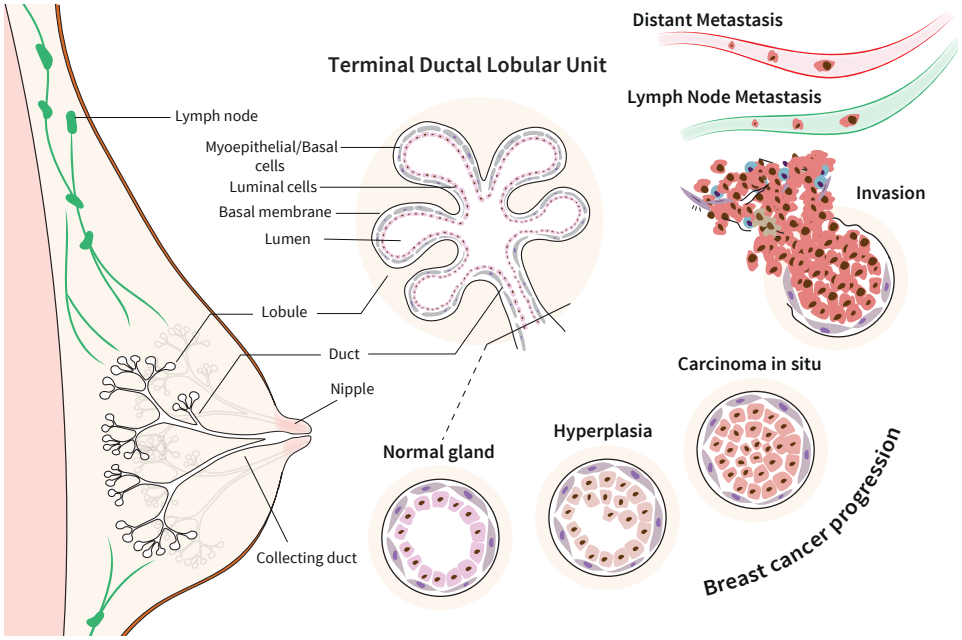


Figure 4.1: Mammary gland structure and breast cancer progression. Breast cancer originates from the bilayer epithelial cells lining the branched system of the mammary gland, often specifically within the milk-producing terminal ductal lobular unit (TDLU). Initially, these cancer cells divide and fill the lumen of the ducts and lobules, a phase termed carcinoma in situ. Subsequently, these cells may penetrate the basement membrane, invading adjacent tissues in a stage called invasive carcinoma. Eventually, the cancer cells gain the potential to enter the bloodstream and metastasise to distant body sites.

opment progresses, unipotent basal and luminal progenitors emerge [Rios et al. 2014; Tao et al. 2014; Van Keymeulen et al. 2011]. Regulators such as the *TP63* gene are suspected to stabilise basal cell identity, while the *NOTCH1* gene probably solidifies luminal identity, at least in mice models [Spina et al. 2021]. This hierarchical pattern, commonly observed in various tissues, tends to mitigate the risk of cancer development as discussed in Section 1.2.

During puberty, sex hormones such as oestrogen drive ductal elongation, branching in the mammary gland, and the formation of milk-producing alveolar expansions grouped in terminal ductal lobular unit (TDLU) (Figure 4.1). Although only a subset of luminal cells express oestrogen receptor (ER+), these cells effectively transmit signals across the entire epithelium. This signalling occurs in a paracrine manner and involves interactions with breast stroma. In murine models, oestrogen stimulates the release of epidermal growth factors (EGFs).

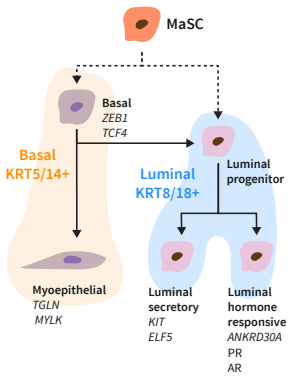
These EGFs bind to their corresponding receptors on stromal cells, initiating the expression of fibroblast growth factors (FGFs). FGFs, in turn, promote the proliferation of luminal cells. Additionally, stromal cells secrete insulin-like growth factor 1 (IGF1), further stimulating mammary gland morphogenesis [Macias et al. 2012].

Breast development extends beyond puberty, undergoing substantial remodelling and increased branching during pregnancy. This transformation, particularly in epithelial and alveolar structures, is triggered by hormones such as progesterone and prolactin, which prepare the gland for lactation. Following weaning, many glandular structures involute, leading to the apoptosis of secretory cells [Macias et al. 2012].

These observations imply the existence of persistent progenitor cells within the tissue or a general capacity for cell proliferation. Such mechanisms are essential, as a significant portion of mammary tissue must regenerate to accommodate subsequent pregnancies.

Research by Prater et al. [2014] indicates that up to 65% of basal cells can form ductal-lobular structures *in vivo*. Single-cell transcriptional trajectory analysis reveals that a substantial fraction of KRT14$^+$ basal cells act as a transcriptional pivot point between luminal and specialised myoepithelial cells [Q. H. Nguyen et al. 2018]. The equipotency of basal cells in the mammary gland is noteworthy, especially given its implication in elevated cancer risk, as discussed in Section 1.2.

One plausible explanation for this characteristic could be the evolutionary necessity for rapid mammary gland remodelling with each successive pregnancy. As mammals evolved, the imperative for tissue renewal may have outweighed the need for cancer protection. This evolutionary trade-off might now manifest as an elevated risk of breast cancer as mentioned in Section 4.1.



Continuous lineage hierarchy from Q. H. Nguyen et al. [2018].

4.1.2. Breast cancer staging and survival

Most breast cancers are adenocarcinomas originating from the epithelial cells in the breast's ductal system, frequently from TDLUs.[1] In the early stages of cancer development, genetically altered precancerous cells grow uncontrollably, a process known as hyperplasia. These cells may also display atypical features. Such aberrant growth often fills the lumens of ducts and lobules, a condition described

[1] Surprisingly, many scientific papers and even the Dana-Farber Cancer Institute's website identify the TDLU as the **sole site of origin for all breast cancers**. This assertion appears unlikely, given that any dividing cell has the potential to become cancerous. The study by Tabár et al. [2014] suggest that 25% of cases actually originate in the duct and show no link to TDLU.

as carcinoma *in situ* (CIS). While CIS can remain confined to the ducts throughout an individual's lifetime, in some cases these cells breach the basement membrane, invading the surrounding stroma. This invasive behaviour marks the transition to invasive carcinoma. At this stage, some cancers can develop the capability to enter the bloodstream or lymphatic system, and form metastatic deposits outside of the breast. Once distant metastases are established, a breast cancer is invariably incurable (Figure 4.1).

Anatomical staging

In managing a disease with such diverse clinical outcomes, clinicians use classification systems to select the most appropriate treatment strategies. The vast majority of primary breast cancers will be completely excised by surgery. The threat for most breast cancer patients is therefore an invisible one – the risk that cancer cells have already spread to distant sites and will reawaken to cause metastasis at a later date. The major challenges for an oncologist are therefore to establish a) whether microscopic metastatic disease is likely to be present and b) which treatment strategies are most likely to eliminate or control it as opposed to simply causing unnecessary side-effects. The most widely used and well validated strategies for achieving this aim draw on both pathological features and anatomical staging of the cancer.

The TNM staging system is used worldwide and helps to classify breast cancers based on the anatomical extent of the disease: the size and reach of the primary tumour (T), lymph node involvement (N), and the occurrence of distant organ metastases (M). This anatomical staging is defined at first diagnosis 'clinically' through physical examination and radiological imaging. Following primary surgery, macro and microscopic analysis of the excised tumour permits detailed 'pathological' staging. In general, larger cancers and/or those with extensive spread to the regional lymph nodes have a worse prognosis and this reflects the higher incidence of micrometastatic disease at diagnosis.

Histological and molecular classification

In global clinical practice, the TNM classification system serves as a cornerstone. However, the system has limitations; specifically, it inadequately captures the biological heterogeneity of breast cancers across different patients. Therefore, additional methods are employed to complement the TNM system to facilitate a personalised approach to cancer management and prognostication.

Histology provides valuable insights into the level of dedifferentiation and proliferative activity of breast cancer, similar to other cancer types, as discussed in Section Section 1.2. Tumour grade, ranging from 1 to 3, encapsulates this information. Grade 3 indicates the most aggressive tumours, often comprising less differentiated cells, while grade 1 represents the least aggressive forms [Elston et al. 1991; Greenough 1925]. Features like tumour grade and the presence of lymphovascular invasion help to stratify those who are most likely to benefit from chemotherapy.

The majority of breast cancer cases are histologically classified as either No Special Type (NST) (~80%) or 'lobular' (~10%).

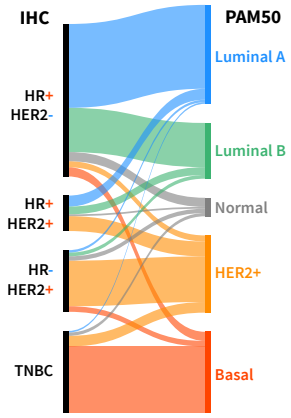
In molecular diagnostics for breast cancer, the standard protocol includes IHC classification for three key markers: oestrogen receptor (ER), progesterone receptor (PgR), and human epidermal growth factor receptor 2 (HER2).

Treatments aimed at lowering oestrogen exposure are effective for all **Hormone Receptor (HR)** positive breast cancers, which express either ER or PgR. These treatments, such as Tamoxifen, aromatase inhibitors, and ovarian suppression, have been proven to improve overall survival, disease free survival and reduce the both local and distant metastasis [Early Breast Cancer Trialists' Collaborative Group (EBCTCG) 2005].

HER2 is another important marker, which is a receptor tyrosine kinase and a common breast cancer oncogene (Figure 4.2). The high level of expression of HER2 is associated with a higher levels of dedifferentiation and proliferation. Development of HER2-targeted therapies (e.g. Trastuzumab) has significantly improved the survival of patients with HER2-positive tumours [Slamon et al. 2001].

Based on the aforementioned molecular markers, cancers are categorised into three distinct clinical subtypes : HR+/HER2- (luminal); HER2+ (HER2) and HR-/HER2- (TNBC). The HR+/HER2- subtype, long recognised as biologically heterogeneous, presents a clinical dilemma in balancing treatment benefit while avoiding unnecessary toxicity. While many HR+/HER2- cancers are cured with surgery and endocrine therapy, others follow a highly aggressive course that may require a more intensive chemotherapeutic strategy.

In routine clinical practice, one common approach for stratifying HR+/HER2- cases involves evaluating their expression of the proliferative marker Ki67, measured via IHC. These classifications roughly correspond to the Luminal A (low Ki67) and Luminal B (high Ki67) intrinsic subtypes as defined by modern RNA-based assays (Table 4.1).



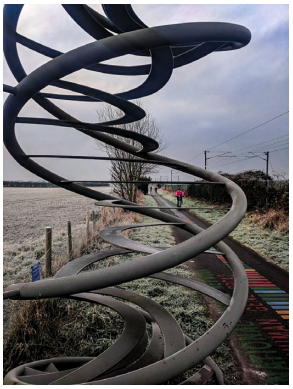
Although gene expression and protein-based systems contain similar clusters, the alignment between them is not precise. Data from Kim et al. [2019].

Table 4.1: Summary of breast cancer bioloigcal subtypes. Overview of the general characteristics of common breast cancer subtypes defined my biomarkers.

Subtype	IHC Status	General Characteristics
Luminal A	HR+ HER2- Ki67↓	Generally lower-grade cancers, have a better prognosis compared to other subtypes.
Luminal B	HR+ HER2- Ki67↑	Often higher-grade than Luminal A, generally more aggressive and less responsive to hormone therapy.
HER2	HR+/- HER2+	Generally more aggressive and treated with targeted therapies.
TNBC	HR- HER2-	Generally higher grade, more aggressive, and poor prognosis.

The advent of high-throughput technologies, including microarrays and next-generation sequencing, has facilitated the creation of classification systems like PAM50 [Parker et al. 2009], MammaPrint ['t Veer et al. 2002], Oncotype Dx [Paik et al. 2004], and IntClust [Curtis et al. 2012]. These systems can offer a more granular view of breast cancer's biological heterogeneity. While largely confirming existing subtype clusters, they further divide them into subclasses with distinct biological characteristics and survival outcomes. Despite their analytical potential, these advanced systems have not been universally adopted in clinical practice for various reasons including economic cost and the scarcity of validation data from prospective clinical trials at the time of writing.

Future improvements likely require a deeper understanding of the disease, including an appreciation for both intra- and inter-tumour heterogeneity. A complexity which arises from the interplay between the evolving genomic landscape of cancer cells and the tissue microenvironment that influences how the tumour develops (Chapter 1).



Cycling path in Cambridge[2] with the entire sequence of BRCA2.

4.1.3. Genomic landscape

Pioneering studies by Nowell et al. [1960] and Stehelin et al. [1976], established a critical link between genomic aberrations and cancer, thereby integrating evolutionary theory with our understanding of cancer as a disease [J. Cairns 1975; Nowell 1976]. In the following decades, many cancer genes such as *TP53* [Lane et al. 1979], *MYC* [Kohl et al. 1984], *BRCA1/2* (on 13q12-13) [Hall et al. 1990; Wooster et al. 1994] [Hall et al. 1990; Wooster et al. 1994] , and *PIK3CA* [Chang et al. 1997] have been identified. Post 2000, high throughput genomic sequencing enabled large-scale, systematic genomics analyses of thousands of tumours. These stud-

ies have revealed complex patterns of genetic alterations.

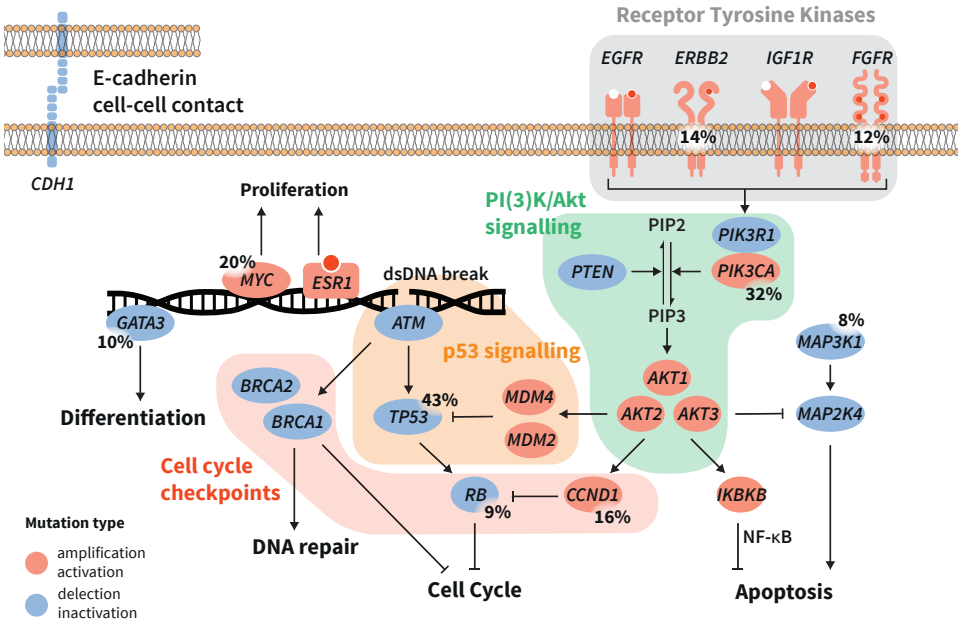


Figure 4.2: Main driver genes and signalling pathways in breast cancer. Comprehensive early genomics studies have identified key driver genes associated with breast cancer. These genes primarily target four main pathways: Receptor Tyrosine Kinases, PI(3)K/Akt, p53 signalling, and cell cycle checkpoint regulators. Mutations in genes contolling these pathways results in: cell survival through inhibition of apoptosis, proliferation, dedifferentiation and faults in DNA reparation. On the figure red indicates gain-of-function mutations or amplifications; blue represents loss-of-function mutations or deletions.

Breast cancer driver genes

Several genomics studies of breast cancer identified thousands of likely driver mutations, spanning hundreds of cancer-related genes and several non-coding regions [Banerji et al. 2012; Cancer Genome Atlas Network 2012; Ciriello et al. 2015; Curtis et al. 2012; Nik-Zainal et al. 2016; Pereira et al. 2016; Shah et al. 2009; 2012; Stephens et al. 2012]. The most commonly mutated breast cancer genes and gene regions as found in Cancer Genome Atlas Network [2012] are

[2] My processing speed is approximately 60 bp/s, as I can slide along 10,257 base pairs of *BRCA2* sequence in 2 minutes and 47 seconds. For comparison, *Taq* polymerase operates at a rate of 150 bp/s, while the high-fidelity *Pfu* polymerase works at a considerably slower pace of 15 bp/s. I deeply thank Lukas Weilguny for consistently providing slipstream support along the way.

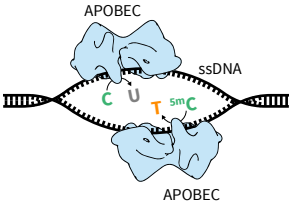
TP53 (43%), *PIK3CA* (32%), *MYC* (20%), *CCND1* (16%), *PTEN* (16%), *ERBB2* (14%), Chr8:(*ZNF703*/*FGFR1*) (12%), *GATA3* (10%), *RB1* (9%), *MAP3K1* (8%). These genes predominantly participate in cell cycle regulation, reparation and growth factor signalling (Figure 4.2). However, most mutations in cancer-associated genes are infrequent, leading to significant genetic diversity among patients.

The mechanisms through which these driver genes are generated vary significantly. While some mutations are inherited, most arise somatically (Chapter 1), frequently as a result of both mutagenic factors and imperfections of the repair systems. High-throughput genome sequencing, especially WGS, has not only facilitated the identification of driver genes but also elucidated the complex mutational processes responsible for their occurrence.

Mutational processes

While DNA microarrays and exome sequencing have identified broad copy number profiles and mutational landscapes in genes, their coverage remains limited and biased. The advent of massively parallel WGS overcame these limitations by providing comprehensive data, allowing for detailed analysis across entire genomes. This advance enabled the decomposition of the mutational landscape into additive sets of mutational signatures, many of which have clear biological implications [Nik-Zainal, Alexandrov et al. 2012].

Breast cancer mutagenesis involves a variety of mechanisms that generate a spectrum of mutation types. Chromosomal rearrangements frequently occur and can be linked to periods of chromosomal instability [Curtis et al. 2012; Gerstung et al. 2020; Nik-Zainal, Alexandrov et al. 2012]. When homologous recombination-based DNA damage repair is defective, often due to inactivating mutations in *BRCA1* or *BRCA2*, error-prone non-homologous DNA end-joining activity emerges. This leads to unique mutational signatures, including both point mutations and a high number of tandem duplications and deletions, particularly in regions of microhomology.



APOBEC is a cytidine deaminase. Under normal conditions it serves as a cellular defence against viral infections by mutating viral ssRNA and ssDNA. In cancer, dysregulated APOBEC mutates ssDNA in the nucleus, causing a high frequency of C>T mutations.

These rearrangements are often associated with mutational signatures related to BRCA1/2 deficiency and other DNA double-strand break repair pathways [Nik-Zainal, Alexandrov et al. 2012].

Another common cause of hypermutation in around half of breast cancers is APOBEC enzyme mutagenesis [Banerji et al. 2012; Nik-Zainal, Alexandrov et al. 2012; Nik-Zainal et al. 2016]. The APOBEC signature consists of a high frequency of C>T and C>G mutations in the TpC context, sometimes exhibiting focal hypermutation patterns, known as kataegis, that frequently coincide with structural

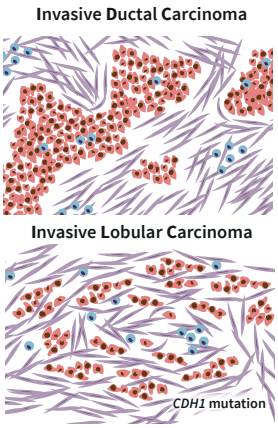
variant breakpoints. Given the diversity of driver genes and the extended period of genome instability, breast cancer appears to be driven by a cumulative effect of multiple accumulated drivers.

Somatic mutations in breast cancer subtypes

Notably, the frequency of driver mutations and mutation signatures varies among breast tumour types. The **Luminal class** commonly features somatic mutations that activate the PI3K-AKT signalling pathway (*PIK3CA*, *PTEN*) and inactivate the GATA3 and JUN kinase (*MAP3K1*) pathways. Tumour suppressors *TP53* and *RB1* are less frequently implicated in Luminal A cancers but are often inactivated in the more aggressive Luminal B subtype [Cancer Genome Atlas Network 2012].

The **TNBC class** aligns closely with the basal-like intrinsic expression subtype, with approximately 75% of TNBCs being basal-like [Shah et al. 2012]. This class exhibits a high frequency of *TP53* mutations, especially among basal-like types, as well as a high level of CNAs [Cancer Genome Atlas Network 2012; Shah et al. 2012]. This coincides with frequent inactivation of genes involved in double-strand break reparation such as *BRCA1* and *BRCA2*, more so than in other subtypes. However, the mutation landscape of TNBC is highly diverse overall [Shah et al. 2012].

In the **HER2+ class**, there is frequent high copy number amplification of *ERBB2*, which accounts for elevated levels of HER2 protein expression. This amplification often co-occurs with amplification of the nearby gene, *GRB7*. Activating hotspot mutations in the *ERBB2* gene can contribute to the overstimulation of HER2 signalling. A systematic analysis of this group revealed subclusters that resemble luminal and basal types in both gene expression and mutation distribution. This suggests that rather than constituting an entirely independent intrinsic type, this group represents a composite of the broader breast cancer spectrum, further complicated by *ERBB2* amplification [Ferrari et al. 2016].



Ductal and lobular breast cancer are two histological subtypes often differentiated by the presence of a single mutation in the CDH1 gene.

While histological subtypes generally exert limited influence on treatment stratification, certain 'special' subtypes exhibit distinct molecular aberrations. These aberrations can account for unique clinical presentations.

Invasive **lobular** breast cancer, the most prevalent 'special' histological subtype, frequently harbours *CDH1* mutations, observed in 45-95% of cases depending on the methodology and cohort [Ciriello et al. 2015; McCart Reed et al. 2021]. This gene encodes E-cadherin, an essential component of adherens junctions that facilitate cell-cell adhesion. The compromised function or loss of E-cadherin is considered a driving factor for the lobular phenotype. Recent studies reveal that

this phenotype correlates with unique cancer behaviours. Although most lobular cancers belong to the Luminal type, they demonstrate a poorer long-term prognosis compared to other Luminal cancers. Moreover, these cancers metastasise to unconventional sites, including gastrointestinal and gynaecological regions, beyond the common locations of liver, lung, and bone [McCart Reed et al. 2021].

Overall, bulk genomic studies indicate that the breast cancer genome exhibits significant inter-tumour diversity. However, while some evolutionary patterns emerge across patients, the presence of mixed phenotypes and observed subclonal diversity within a single tumour complicates our understanding of tumour evolution [Ciriello et al. 2015; Pereira et al. 2016; Shah et al. 2012]. Initial bulk studies on large cohorts yielded basic insights into intra-tumour heterogeneity, but these were largely confined to broad statements of diversity. A nuanced understanding of evolutionary history at the subclonal level within a single tumour is crucial for comprehending both the biology and therapeutic responses of patients.

4.1.4. Clonal evolution

Progress in WGS, multi-regional tissue sampling techniques, and the development of advanced computational tools collectively enable more intricate studies of breast cancer evolution (Boxes 1 and 3). The first study by Nik-Zainal, Van Loo et al. [2012] used WGS data from single samples complemented with bespoke computational methods to reconstruct the evolutionary history of 21 breast tumours in detail. The study traced the tumour's lineage back to the zygote, delineating multiple clonal branches and establishing the temporal sequence of chromosomal rearrangements and driver mutations. Notably, each studied tumour was found to contain a dominant subclonal lineage, making up over half of the cancer cells. The researchers proposed a model wherein sparse, long-lived cellular lineages accumulate mutations passively until a critical event triggers their rapid proliferation, culminating in a clinically detectable tumour mass.

Other pioneering studies relied on CNA inferred from single-cell sequencing that allowed to reconstruct phylogenies. They also showed that breast cancer is a patchwork of clones emerging from 'punctuated' clonal expansions, rather than linear progression [Gao et al. 2016; Navin et al. 2011; Y. Wang et al. 2014]. Furthermore, these works indicated that clonal expansions often correlate with focal CNA events. Although single-cell sequencing enhanced our understanding of divergence beyond major expansion events, the technique has limitations. These included low sensitivity and challenges in the accurate reconstruction of phylogenetic trees based on copy number profiles alone (Box 3).

Box 3: Tumour Life History rectionstruction from WGS data

Remarkably, deep WGS data provide sufficient information to partially reconstruct the life history of the tumour cell population, from zygote formation to the day of tumour sampling [Dentro et al. 2017; Nik-Zainal, Van Loo et al. 2012]. This is possible because mutation distribution contains information about the population structure that could be modelled computationally.

Somatic mutation calling

The initial step entails differentiating genuine single nucleotide variants (SNVs) originating from cancer cells from normal germline cells and sequencing errors. The procedure can also extend to structural variants, which involves identifying corresponding sections of sequence reads. However, this task becomes more challenging when dealing with short-read sequencing.

Cancer Cell Fraction

For bulk genomic sequence data, the VAF for each mutation is calculated using read information:

$$VAF = \frac{r_{mutant}}{r_{total}} \tag{4.1}$$

VAF alone, however, fails to capture the tumour's subclonal architecture due to influences like tumour purity ρ and local copy numbers $n_{loc,t}$ and $n_{loc,n}$. Instead, mutation copy number n_i is used:

$$n_i = \frac{VAF_i}{\rho}[\rho n_{loc,t} + (1-\rho)n_{loc,n}] \tag{4.2}$$

Algorithms such as Battenberg provide values for (n_{loc}) and (ρ), by partitioning the genome into distinct copy-number profiles.

In instances of multiple alleles or subclonal copy number changes, a complex model typically is necessary to accurately infer the number of chromosomes bearing the mutation n_{chr}. The CCF is then calculated as:

$$CCF_i = n_i / n_{chr} \tag{4.3}$$

Note, that a fully clonal mutation will have a CCF of 1

Subclonal architecture inference from SNVs

Clonal expansions within the tumour elevate the frequencies of lineage-specific mutations. Overlooking sporadic mutations that occur during this expansion allows us to model the data as a mixture of binomials. A Dirichlet process serves well for the task of inference of unknown number of clusters. The model is as follows:

$$r_i \sim Binomial(r_{total,i}, \zeta_i \pi_i) \tag{4.4}$$

$$\pi_i \sim \mathcal{DP}(\alpha, P_0) \tag{4.5}$$

Here ζ_i the proportion of reads expected for a fully clonal mutation, and π_i is the true fraction of tumour cells carrying the mutation. π_i is drawn from a stick-breaking process with the base distribution P_0 (locations of clusters) and concentration parameter α (the probability weights).

Subclonal architecture inference from CNAs

CNAs can also inform subclonal population structures. These alterations are estimated by read depth and the disparity between maternal n_A and paternal n_B allelic reads. B-allele frequencies (BAF) indicate these changes:

$$BAF_i = \frac{n_{B,i}}{n_{A,i} + n_{B,i}} \tag{4.6}$$

Deviations from expected BAF, adjusted for tumour purity, point to subclonal copy number changes. Usually, at least two subclonal populations differing by a single copy number change must be identified.

Phylogenetic tree reconstruction

The "pigeonhole principle" commonly applies to discern evolutionary relationships among subclones. It says that the sum of CCFs for subclones

cannot exceed that of their ancestor. Trees assemble accordingly. But pigeonhole alone can still lead to multiple tree solutions, e.g. in the case of a 30% and 50% subclone.

Multiregional sequencing additionally provides better resolution of subclones with similar CCFs.

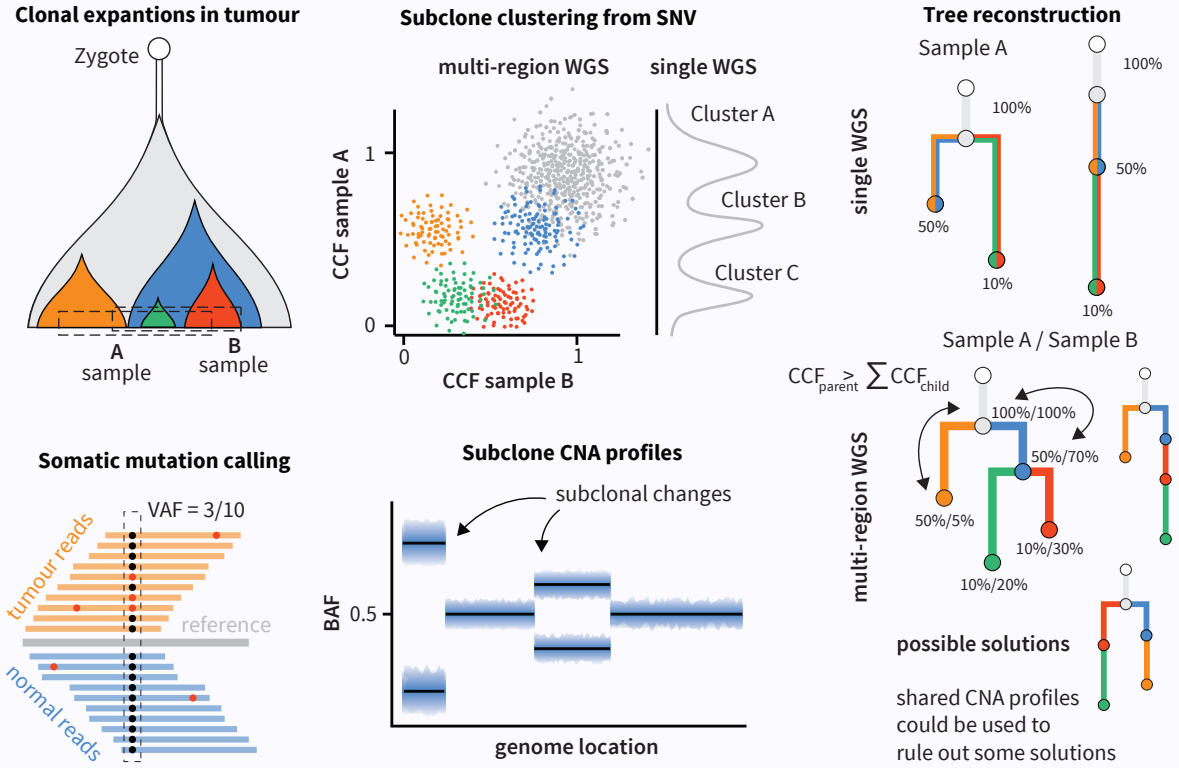
Beyond single sample WGS

Conducting the analysis in a **multi-regional** way enhances resolution. In a single sample, subclonal clusters may coincidentally exhibit similar CCFs. However, these clusters are likely to display different proportions in another sample. This variability enables the resolution of smaller clusters with lower presence. The natural extension of this strategy is **LCM-WGS** (Box 1).

Single-cell WGS offers significant improvements in inferring subclonal architecture, particularly in enhancing the resolution of the phylogenetic tree's leaves. This technique has a clear advantage over bulk WGS in its ability to identify changes present in a single cell that has not yet undergone clonal expansion. However, while single-cell WGS often provides enough coverage for reconstructing CNA profiles, it faces challenges in accurately constructing phylogenetic trees based on SNVs as the data is sparse. Moreover, although tree inference based solely on CNAs is feasible, the results are typically even more ambiguous due to a higher volatility of CNAs compared to SNVs, making it difficult to draw definitive conclusions about evolutionary dynamics.

Timing of evolutionary branches

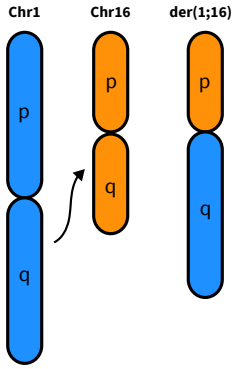
Mutations are generally irreversible and increase monotonically over time, offering reliable event ordering within the same branch. However, they do not guarantee chronological timing due to variable mutation rates. To improve chronological estimates, one can limit the analysis to mutational processes ubiquitous in normal cells, like SBS1 or SBS5, and incorporate the patient's age. This method also enables timing of tumour genome rearrangements, as outlined in Gerstung et al. [2020].



Yates et al. [2015] conducted the first multi-regional WGS study of primary breast cancers. Although the study's sensitivity was not exceptionally high, its use of a multi-region approach enhanced branch resolution over earlier studies, unveiling considerable genomic diversity across samples. Most cases exhibited long trunk mutations, similar to Nik-Zainal, Van Loo et al. [2012], but the study also reported instances of early branching and extensive parallel evolution. Notably, mutations in the *PTEN* and *TP53* genes showed signs of convergent evolution, occurring independently across parallel branches. This study further confirmed that chromosomal events can happen at any evolutionary stage, both early and late, reshaping the genome and contributing to subclonal diversity.

Enhanced spatial resolution in sampling strategies, facilitated by techniques like tissue microdissection, enables more precise inquiries into breast cancer development (Box 1). In the study by Casasent et al. [2018], LCM isolated cells from both ductal and adjacent invasive regions of ductal carcinomas. Contrary to previous assumptions about breast cancer invasion, identical clones were found both inside and outside the ducts without significant changes. This suggests that the evolutionary changes necessary for invasion occur within the ducts and are initially contained by them. Another microdissection study paired with single-cell DNA sequencing by Lips et al. [2022] concentrated on invasive recurrence. The study linked this recurrence to untreated clones that had been present for an extended period within DCIS.

The more comprehensive the microdissection strategy and study depth, the greater the number of parallel branches observed in the reconstructed phylogenetic trees. A detailed study by [Nishimura et al. 2023] focused on the early evolution of DCIS. The research revealed that many breast cancers exhibited a recurring chromosomal abnormality, der(1;16), acquired early in cancer evolution, roughly around puberty. These der(1;16)-positive clones expanded across large breast regions, spawning multiple independent cancer branches as well as non-cancerous lesions (Figure 4.3). Conversely, most der(1;16)-negative clones remained restricted to single lobules post-puberty. The findings imply that a branching, multifocal model of cancer progression might be more prevalent than linear models. Additionally, the study revealed no correlation between the number of driver events and histology. This suggests that local microenvironments or other unmeasured intrinsic factors may play a role in cancer development. Further discussion on this discrepancy appears in Section 5.4.



Unbalanced translocation between Chr1 and Chr16 was frequently observed in study by Nishimura et al. [2023].

Overall, the studies discussed above portray cancer evolution as a complex, multifaceted process characterised by considerable intratumour heterogeneity on a genomic level. Earlier studies focused on a most dominant clone reporting a

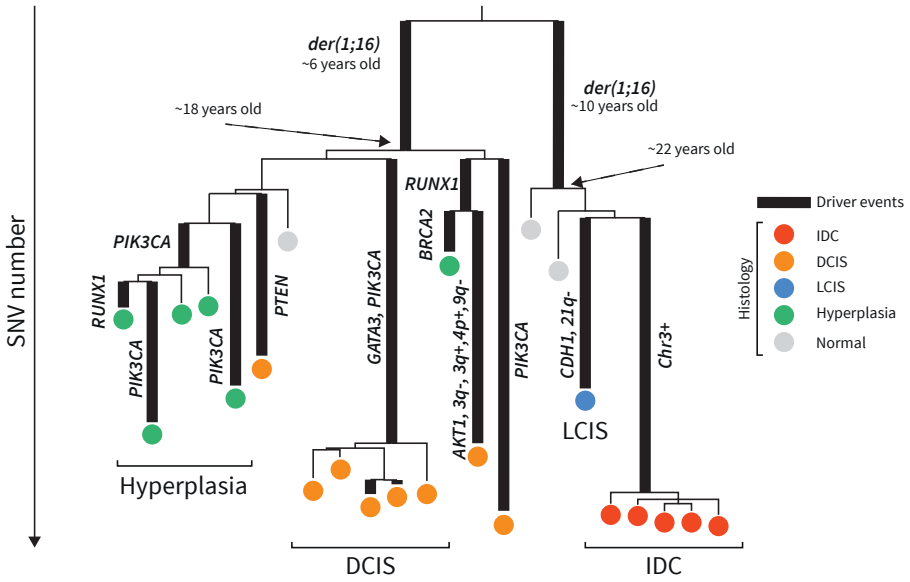


Figure 4.3: Highly branched parallel evolution of breast cancer described by Nishimura et al. [2023]. Deep study of breast cancer case reveals complex, highly branched evolutionary path. Key driver events, specifically der(1;16) translocations, emerged independently in two branches when the patient was around 6 and 10 years old. Although only one branch led to an invasive phenotype, parallel driver events occurred across multiple branches. The complex relationship between genome and phenotype becomes evident from the mismatch between the number of driver events and the resulting phenotype. This suggests a potential role for the microenvironment in disease progression.

mostly linear progression of cancer, likely due to technological limitations and undersampling. In contrast, more recent research uncovers a complex landscape with multiple parallel branches that might diverge early in cancer evolution. The mechanisms driving these branches remain a topic of debate, with possibilities including both positive selection pressure and neutral drift [Sottoriva et al. 2015; M. J. Williams et al. 2016]. The relationship between driver events and specific histology also remains ambiguous, even when focusing on individual evolutionary branches. Likely, this limitation arises because genomic studies focus solely on the intrinsic properties of cancer cells. While the extrinsic microenvironment probably plays but as yet poorly understood role in shaping cancer evolution.

4.1.5. Microenvironment

Single-cell proteomics and transcriptomics studies have substantially advanced our understanding of the tumour microenvironment in breast cancer, going well beyond bulk studies like PAM50 transcriptional classification [Pal et al. 2021; Wagner et al. 2019; S. Z. Wu et al. 2021]. These investigations confirm that the general structure of the breast cancer microenvironment closely aligns with the broader cancer landscape, as outlined in Section 1.3. The microenvironment commonly comprises immune cells (both myeloid and lymphoid), stromal cells (including cancer-associated fibroblasts (CAFs), endothelial cells, perivascular-like cells (PVLs) and adipocytes), and epithelial cells (luminal and myoepithelial). Breast cancer exhibits significant patient-to-patient variation in the relative proportions of these cell types. Moreover, spatially organised recurrent cellular communities correlate with patient survival [Danenberg et al. 2022; Jackson et al. 2020].

Immune cells

In the context of cancer, immune cells are often simplistically categorised as either tumour-promoting, such as regulatory T cells (Tregs) and M2 macrophages, or tumour-inhibiting, such as cytotoxic CD8+ T cells, natural killer (NK) cells, CD4+ T-helper cells, and M1 macrophages (Figure 1.4). Generally, the immune system acts as a tumour suppressor, and the presence of tumour infiltrating lymphocytes is linked to a better prognosis (Figure 1.4). However, lymphocytes can enter a so-called 'exhausted' dysfunctional state, characterised by the expression of co-inhibitory receptors such as CTLA-4, PD-1 and LAG3 confirming that this is likely an oversimplification.

Indeed the landscape of immune cell roles in cancer is likely much more complex. For instance, while NK cells generally target tumour cells, some studies suggest they might also enhance tumour vascularisation and even exert immunosuppressive effects [Retecki et al. 2021]. Furthermore, T-cell activation states in breast cancer have been reported to represent a continuous and complex spectrum rather than the discreet categories reported in normal breast [Azizi et al. 2018; Savas et al. 2018]. T-cell diversity largely stems from variations in T-cell receptor (TCR) and antigen-presenting cell heterogeneity [Azizi et al. 2018]. Such variability may result from differing tumour antigens across clonal populations. Such variability may result from differing tumour antigens across clonal populations. Finally, while myeloid cells are generally thought to display quite distinct

activation states it has recently been reported that a single myeloid cell can simultaneously express both M1 and M2 programs [Azizi et al. 2018].

Spatial analysis of immune cell distribution in breast cancer indicates non-uniform patterns linked to prognosis [Danenberg et al. 2022; X. Q. Wang et al. 2023]. For instance, dysfunctional T cells often cluster near Tregs, which are believed to control their proliferation and activation. Large dysfunctional T-cell aggregates may signify tumour cells resistant to ongoing immune attacks, thereby chronically stimulating cytotoxic T cells. Conversely, structures containing antigen-presenting cells, macrophages, and T cells associate negatively with survival. In contrast, cell communities featuring granulocytes, and particularly tertiary lymphoid structure (TLS) with B cells, correlate with better survival outcomes.

Stroma

In the stroma, fibroblasts are the most extensively studied cells. Researchers typically categorise these fibroblasts into two broad types: SMA- fibroblasts and SMA+ myofibroblasts [Costa et al. 2018]. While efforts exist to categorise more types, these efforts face challenges due to the cell-type fluidity of fibroblasts [Cords et al. 2023; S. Z. Wu et al. 2021]. These cells serve essential structural functions and engage in immune cell recruitment and extracellular matrix remodelling.

Of specific interest is the SMA+ FAP+ subset of fibroblasts, which acts as a crucial immunosuppressive agent. This subset attracts T lymphocytes and facilitates their differentiation into immunosuppressive Tregs, thereby aiding in immune evasion in cancer. In contrast, SMA+ FAP- fibroblasts lack this immunosuppressive function [Costa et al. 2018]. Spatial distribution analysis indicates that myofibroblasts frequently reside at the tumour-stromal interface in breast cancer, which suggests their role in lymphocytic exclusion [Danenberg et al. 2022]. Although this seems to promote tumour growth, other fibroblasts located near tertiary lymphoid structures (TLS) correlate with a better prognosis [Cords et al. 2023; Danenberg et al. 2022] (Figure 1.5).

In a different spatial study, Risom et al. [2022] demonstrated that 'desmoplastic' stroma, characterised by a high frequency of fibroblasts and intense collagen deposition around ducts, correlates with a favourable prognosis and reduced likelihood of progression to invasive cancer. Notably, this type of stroma coexists with a thin and discontinuous myoepithelial layer. This type of stroma may indicate a robust immune response, facilitated by a permeable myoepithelium that exposes

the cancer cells to immune surveillance.

Endothelial cells and PVLs are other significant components of the tumour microenvironment. These cells constitute the vasculature, critical for tumour growth and metastasis. In the context of breast cancer, a dense vascular stroma correlates with a poor prognosis [Danenberg et al. 2022].

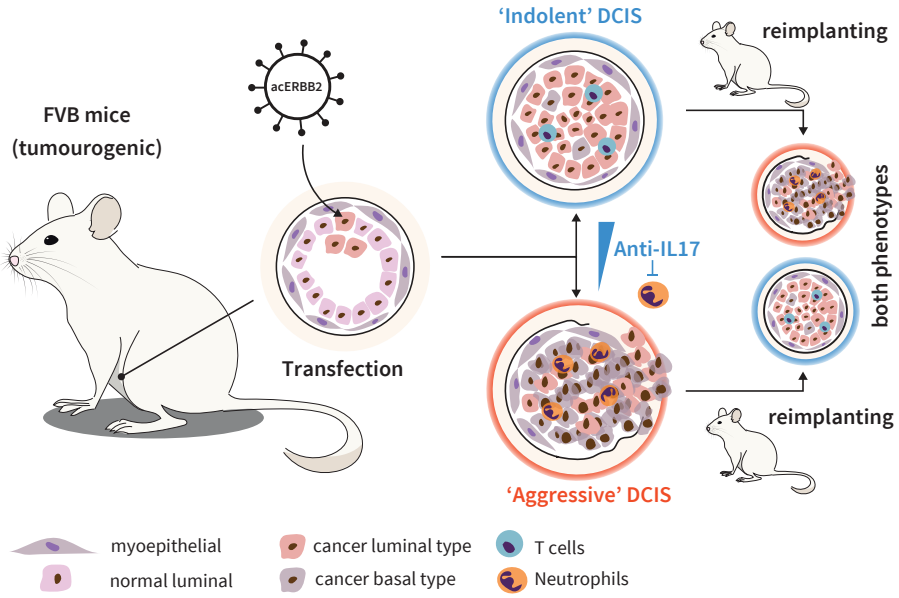


Figure 4.4: Immune niches and breast cancer cell plasticity. Experimental design from Sinha et al. [2021]. A lentivirus vector delivers acERBB2 into mammary gland cells, inducing the formation of ductal carcinoma *in situ* (DCIS). The DCIS cells stochastically adopt one of two phenotypes: 'indolent' or 'aggressive'. The 'indolent' phenotype features a luminal-like state and the presence of T cells, whereas the 'aggressive' phenotype is characterised by a basal-like state and high neutrophil infiltration. Intriguingly, reimplantation of cells from both phenotypes reveals functional indistinguishability, as they produce similar ratios of 'aggressive' and 'indolent' phenotypes in a new host. This experiment suggests that cells with equivalent oncogenic potential can exist in diverse phenotypic states. Additionally, neutrophil inhibition via anti-IL17 treatment reduces the likelihood of the 'aggressive' phenotype emerging, suggesting their role in promoting the 'aggressive' phenotype.

Cancer Epithelial plasticity

As previously discussed, the breast ductal system's epithelium comprises luminal and basal cell types, which exhibit significant plasticity. In cancer, this epithelial plasticity not only persists but also amplifies, leading to a diverse range of transcriptional states [Pal et al. 2021; Wagner et al. 2019; S. Z. Wu et al. 2021]. The

study by S. Z. Wu et al. [2021] dissected the intrinsic expression profiles of breast cancer cells into various transcriptional modules, each showing distinct patterns. Notably, the epithelial-mesenchymal transition (EMT) and proliferative modules appeared mutually exclusive. This observation corroborates the incompatibility between a mesenchymal-like state and proliferation in breast cancer [Tsai et al. 2012]. However, despite the coexistence of multiple tumour cell phenotypes within all tumour ecosystems, one phenotype often dominates, potentially reflecting the expansion of the fittest tumour subclones [Wagner et al. 2019].

The extent to which breast cancer plasticity influences invasion has been explored by Sinha et al. [2021] through a mouse model study (Figure 4.4). In this research, cancer cells sharing the same genetic background — specifically, activated *ERBB2* — gave rise to two distinct types of DCIS lesions within the same animal: 'indolent' and 'aggressive'. Importantly, cells from both lesion types demonstrated equal ability to initiate invasive tumours. The key differences between these two lesion types lay in their intrinsic phenotypes and surrounding immune environments. The 'aggressive' lesions contained more basal-like cells and macrophages, while the 'indolent' lesions were enriched for luminal cells and T cells, hinting that local environmental factors influence the tumour's invasive potential.

4.1.6. Open questions on breast cancer progression

Despite considerable advancements in breast cancer research, our understanding of disease progression and underlying mechanisms remains incomplete. We have identified key driver mutations and broadly understand the mutation processes that lead to their accumulation. We are familiar with the basic composition of the breast cancer microenvironment and the cellular communities typically associated with unfavourable clinical outcomes. Yet, a comprehensive understanding of how genomes and environment interact and influence each other eludes us. Our current models of cancer progression are simplistic, primarily because few studies have rigorously examined the phylogenetic structure of patients' breast cancer and even fewer have simultaneously characterised both genetic and TME components. As a result, several critical questions about cancer progression remain unclear:

What are the cells of origin, and to what extent do they dictate the future trajectory of cancer evolution? Several studies suggest expressional similarity of a cancer subtype to the exact cell lineage. For example HR+ Luminal cancers cells are similar to luminal cell types, while TNBC cells have some characteristics

of basal cells [Skibinski et al. 2015; Taurin et al. 2020] (Section 4.1.1). Tracing the cancer and normal cell lineages together with their intrinsic expression profiles may help to elucidate the cells of origin and their role in cancer progression.

Does the environment play a role in the initial onset, or are the initiating factors solely genetic? The study on phenotypically normal tissue, such as the one conducted by Nishimura et al. [2023], reveals that numerous branches may contain driver events yet not appear malignant. A systematic profiling of the environment in which these branches exist could elucidate the environment's role in the initial onset of cancer.

Why do some clones successfully colonise ducts while others do not? The ductal system serves as the environment where CIS evolves before progressing to invasive stages. Research on CIS has reported clonal heterogeneity [Casasent et al. 2018; Nishimura et al. 2023; Yates et al. 2015] and associated recurrence with minor clones that persist within the ductal system [Lips et al. 2022]. The reasons for this diversity and the role of the environment in this process remain unclear.

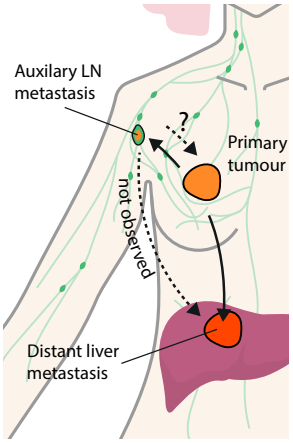
What types of environments trigger invasion, and what genetic backgrounds sustain clones in invasive lesions? Evidence suggests that clones with the same genetic background could exist in both CIS and invasive states simultaneously [Casasent et al. 2018] and that the microenvironment contributes to the onset of invasion [Sinha et al. 2021] (Figure 4.4). However, the extend the genetic backgrounds of these clones influence the process remains unclear. The question of whether any CIS clone can transition to an invasive state, given the right conditions, or if specific genetic backgrounds are required, warrants further investigation.

In the context of metastasis, what roles do the environment and genetics within the primary tumour play, and what is the role of the environment at distant sites? As discussed in Section 1.2, multiple multi-region genomic sequencing studies have explored the phylogenetic links between primary tumours and their corresponding metastases in breast cancer patients [Brown et al. 2017; De Mattos-Arruda et al. 2019; Yates et al. 2015; 2017]. Evidence suggests that breast metastases often arise from clonal, rare seeding events [Brown et al. 2017; De Mattos-Arruda et al. 2019], although De Mattos-Arruda et al. [2019] suggest that polyclonal seeding may be more prevalent. This could be explained by several seeding events, or by an environment favourable for the formation of DTCs clusters, as suggested by mouse models [Cheung et al. 2016]. It is also known that DTCs can seed (pre-)metastatic niches in bone, liver and probably other organs, and here they may remain dormant for years only to be triggered to awaken and cause overt metastasis many years or even decades later. The cause for this may

be cellular intrinsic and/ or a change in the environment [Diamantopoulou et al. 2022; Szczerba et al. 2019]. The challenges associated with detecting and characterising dormant cancer cells in their premetastatic remains a major challenge for cancer research.

Do cancer - immune interactions in the lymph node shape the disease trajectory? Some studies propose that LN metastases indicate cancer cells' ability to survive and grow in other organs, rather than directly aiding their dissemination [Fisher et al. 1977; Ullah et al. 2018]. This is surprising because LN metastases exist in an environment abundant with immune cells. While these metastases may not directly advance to distant organs, they could influence breast cancer progression by modulating the immune response in the primary tumour.

While the spatial genomics technologies outlined in Chapter 2 may not offer immediate, definitive answers to these questions, they do provide a framework for further investigations. Specifically, the BaSISS approach allows for the concurrent profiling of phylogeny and microenvironment across extensive tissue sections. In this chapter, I will show how data generated using this approach could help to investigate the relationship between genetic and microenvironmental evolution in breast cancer progression. Specifically, the analysis traces the evolutionary history of breast cancer from the earliest stages of tumour development to LN metastasis.



While metastasis to the auxiliary LN is common, it doesn't seem to directly relate to the formation of distant metastases. Whether ongoing evolution in the LN influences the primary tumour site is currently unknown.

4.2. Results

4.2.1. Recap of the BaSISS workflow

The BaSISS workflow is centred around fresh frozen tissue blocks that undergo serial cryosectioning to generate tissue for bulk WGS and z-stacked sections for in-tissue spatial clone mapping and spatial phenotyping (Figure 4.5). Following subclone detection from bulk WGS data, there are three core BaSISS steps. First, to facilitate detection of multiple clones of interest, BaSISS padlock probes with sequence-specific oligonucleotide target recognition arms are designed towards both mutant and wild-type alleles of clone-defining somatic variants. A unique 4–5 nucleotide reader barcode on each probe enables multiplexing [Ke et al. 2013]. BaSISS targets can take the form of any expressed somatic mutation, including point mutations and rearrangement breakpoints, and can be supplemented with copy number alterations (Tables A.1 and A.2). Second, BaSISS and transcript detection are performed as previously described for gene expression ISS using cyclical microscopy [Ke et al. 2013; Svedlund et al. 2019] (Figure 4.5

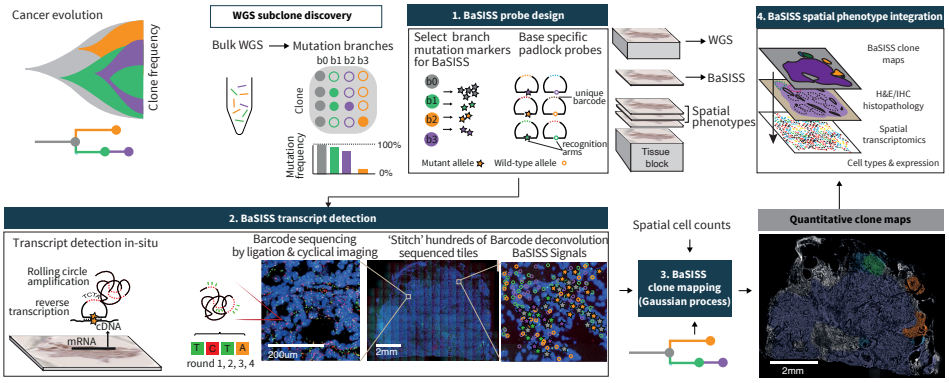


Figure 4.5: The BaSISS workflow to generate cancer clone maps. Following de novo mutation detection and subclone discovery in WGS data, the BaSISS workflow is performed as follows: (1) bespoke mutation-specific padlock probes are designed. (2) BaSISS transcripts are detected. To achieve this, BaSISS padlock probes hybridise to complementary DNA (cDNA) in situ. By virtue of a highly specific DNA ligase, only completely target-complementary padlock probes are ligated and form closed circles. Ligated probes are amplified through rolling circle amplification and their reader barcodes are detected in tissue space through sequencing by ligation with fluorophore-labelled interrogation probes and cyclical microscopy. (3) Mathematical modelling of BaSISS signals and the genotype of clones is then performed to derive clone maps. (4) Subsequent phenotype and microenvironment characterization of clones is then possible, by integrating clone fields with spatial datasets acquired from serial tissue sections. The BaSISS model and cell typing are described in Chapters 2 and 3

and Appendix A.5).

Third, continuous spatial subclone maps are generated using a statistical algorithm that exploits BaSISS signals as well as local cell counts (derived from the DAPI channel during the fluorescence microscopy of BaSISS) using two-dimensional Gaussian processes (Chapter 2 and Figure 2.4). The variational Bayesian model also accounts for unspecific or wrongly decoded signals and variable probe efficiency and is augmented by variant allele fractions in the bulk genomic sequencing data. In an optional, fourth characterisation step, BaSISS clone maps can be aligned and integrated with additional layers of spatial phenotype data. In this study, we performed spatially resolved single-cell transcriptomics using targeted ISS (using a previously published 91 gene oncology, a novel 62 gene immune panel and drawing on published single-cell RNA sequencing data) [Svedlund et al. 2019; S. Z. Wu et al. 2021] and immunohistochemistry (IHC) staining (Chapter 3, Appendices A.2, A.5.4 and A.5.5, and Appendix Figures C.1 and C.2). Additional sections were obtained to perform validation of our workflow using LCM and low-input WGS as previously described [Ellis

et al. 2021] (Section 2.3, Appendix A.7, and Figure 2.6).

4.2.2. Two cases of multifocal breast cancer

The cohort includes eight tissue blocks from two patients (P1 and P2) who underwent a surgical mastectomy for a multifocal breast cancer. These patients were selected to permit a comparison between genetic and histological progression models in early breast cancer development [Cowell et al. 2013] (Figures 4.1 and 4.6). P1 had two separate oestrogen receptor (ER)-positive, human epidermal growth factor receptor 2 (HER2)-negative primary invasive breast cancers (PBCs) within a 5-cm bed of DCIS; we used tissue blocks from both PBCs (samples P1-ER1 and P1-ER2) and three regions from DCIS (samples P1-D1, P1-D2 and P1-D3). P2 had two separate PBCs of the 'triple-negative' subtype (lacking the ER, progesterone receptor and HER2). We sampled both PBCs (samples P2-TN1 and P2-TN2) and an axillary lymph node that contained metastatic cancer deposits (sample P2-LN1) (Figure 4.6).

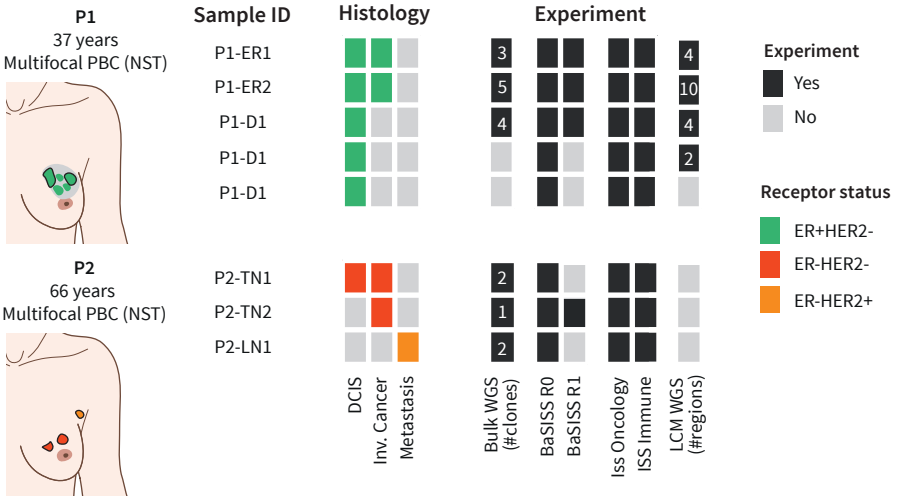


Figure 4.6: Breast cancer cohort description. The two cases of multifocal primary breast cancer (PBC) used to develop the BaSISS approach. Coloured tiles report the histological features within each sample and the experiments performed. The number of clones identified by WGS and targeted by BaSISS are reported as white numerals. Notations are: NST - No special type, ER - oestrogen receptor positive, TN - triple negative, D - ductal carcinoma *in situ*, LN - lymph node.

4.2.3. Charting histogenomic relationships

Histology-driven sampling of well-defined stages of cancer progression can uncover mechanisms and markers of disease progression [Casasent et al. 2018; Cowell et al. 2013; Nirmal et al. 2022; Risom et al. 2022]. Up to two-thirds of PBCs contain both invasive cancer and intermixed DCIS, a non-obligate precursor lesion. How these distinct 'stages' of cancer development might relate to genetic diversification within the same tissue is generally unknown [Kole et al. 2019] (Figure 4.1). To demonstrate that BaSISS can chart these relationships across entire tissue sections, we examined three PBC samples with intermixed invasive and DCIS histology: P1-ER1, P1-ER2 and P2-TN1 (Figures 4.7 and 4.8 and Appendix Figures D.1 and D.2a-c).

BaSISS detected 2–4 subclones per PBC in accordance with bulk WGS data. Clone maps (top panels Figures 4.7 and 4.8) and the quantitative clonal composition of 73 individually annotated microregions (middle panels Figures 4.7 and 4.8, Appendix Figure D.1a,b and Appendix Figure D.2a,b) revealed that individual subclones form spatial patterns that were, by varying degrees, related to the histological progression states. Normal tissue elements, including immune aggregates and histologically normal ducts, appear unstained consistent with a wild-type status for the targeted clones (green and yellow contours, respectively; Figure 4.7). In P1-ER2, an area of hyperplasia was predicted and confirmed by LCM–WGS to be genetically unrelated to the cancer (blue contour; bottom panel Figure 2.6 and top panel Figure 4.7).

In each PBC, the genetic and histological progression models were broadly consistent, in which the invasive disease was mainly composed of cells from the most recently diverged subclone: P1-red, P1-purple and P2-purple in samples P1-ER1, P1-ER2 and P2-TN1, respectively (middle panel Figure 4.7, bottom panel Figure 4.8). By contrast, earlier diverging clones co-localised entirely or in part to the histological pre-invasive lesion: DCIS. For example, in P1-ER2, BaSISS predicted that green branch mutations were completely absent from the invasive compartment, a conclusion that is supported by three separate microdissections (LCM–WGS) from distant regions of invasive cancer in P1-ER2 (bottom panel Figure 2.6 and Appendix Figure D.1c).

However, in each PBC, there was a subclone that spanned both DCIS and invasive histology, revealing that disconnects between histological and genetic progression states can exist. This was the case for clone P1-red in P1-ER1 and clone P1-purple in P1-ER2. These DCIS-invasive spanning clones could be distinguished from each other by hundreds of private mutations, including different

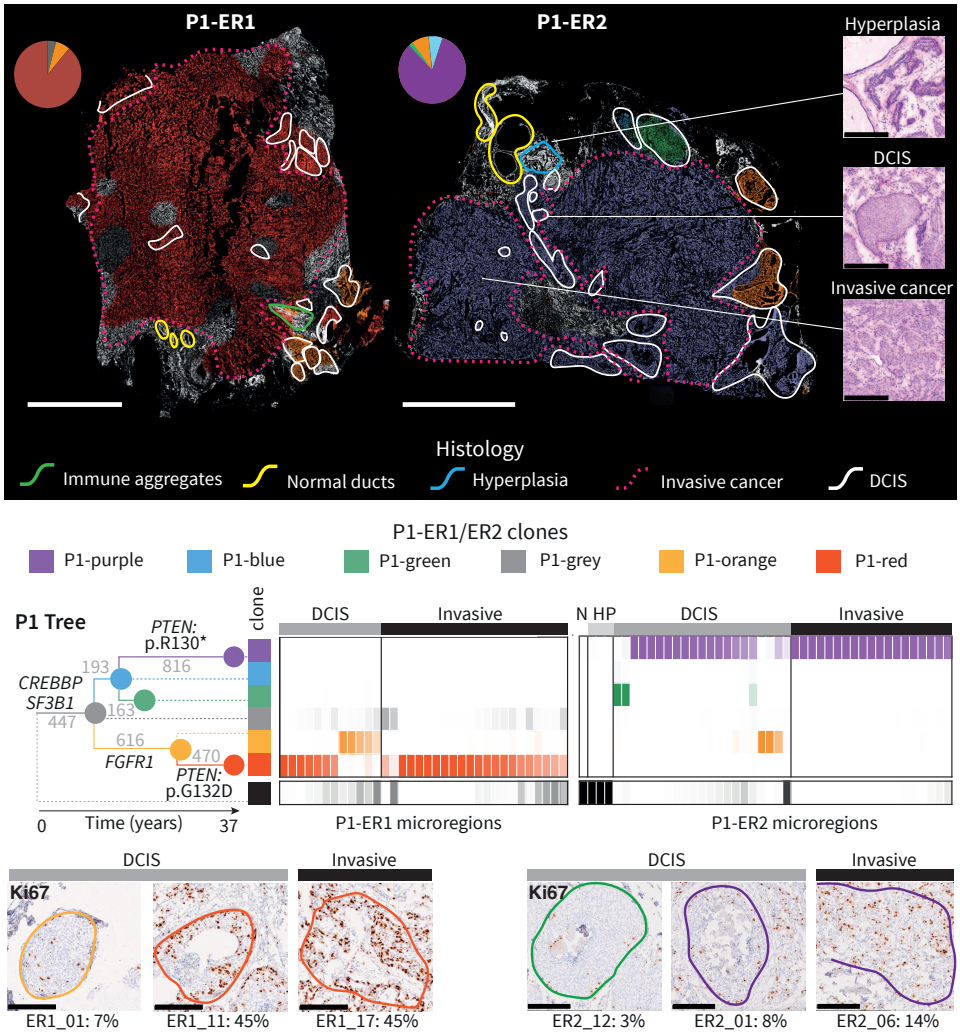


Figure 4.7: Genetic clones mapped in histological context from two PBC samples in P1. ***top***, BaSISS maps of two PBCs from P1 with intermixed DCIS and invasive cancer. The most prevalent genetic clone is projected as a coloured field (corresponds to ***middle*** section) on DAPI images (reported if the CCF > 25% and the inferred local cell density > 300 cells per mm^2). Pie charts report the WGS-estimated clone compositions. Inset images are regions of P1-ER2 (H&E-stained serial tissue sections) that represent three histological progression states. ***middle***, The phylogenetic tree was inferred from P1 multi-region WGS. Branches and nodes are coloured to reflect the clone maps. Heatmaps report clone composition in 34 and 44 histologically annotated epithelial-containing microregions of P1-ER1 and P1-ER2, respectively. Microregions include individual ducts or randomly selected regions of invasive cancer. ***bottom***, IHC per P1-ER1 (left) and P1-ER2 (right) for the proliferative marker Ki-67 in six clone territories (indicated by contour colour); the percentage of nuclei staining positive (brown) is reported. Notations are: HP - hyperplasia; N - normal ducts. Scale bars, 250 μm for microregions and 2.5 cm for clone maps.

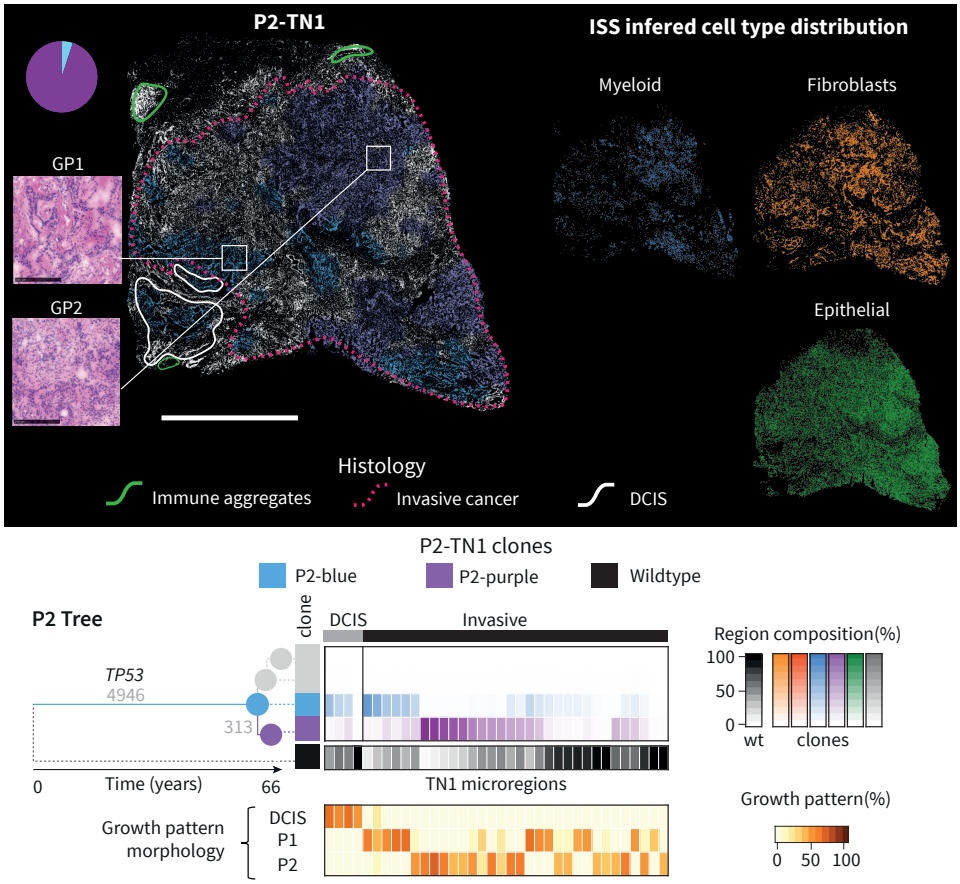


Figure 4.8: Genetic clones mapped in histological context from the PBC sample in P2. ***top***, As in top panel Figure 4.7, but a clone map of P2-TN1. Mini-images report ISS-derived cell types (right) and H&E tissue section snapshots of the two cancer growth patterns (GP1 and GP2) reported in P2-TN1 (left). ***bottom***, Phylogenetic tree for P2 and heatmap of 36 P2-TN1 microregions, as in middle panel Figure 4.7. Branches relating to clones not detected in this sample (that is, only found in P2-LN1) are shaded grey. The bottom heatmap is the estimate by the histopathologist and reports the contribution of different growth patterns to the microregion, defined by distinct nuclear and architectural features. Scale bar, 2.5 mm.

inactivating driver mutations in *PTEN*, indicating parallel evolution along these divergent lineages that resulted in two distinct instances of cancer invasion (total mutation numbers label the phylogenetic tree branches; middle panel Figure 4.7). The spatial predictions of the BaSISS model of intraductal acquisition of *PTEN* mutations and *PTEN* protein loss was confirmed by LCM–WGS and IHC, respectively (bottom panel Figure 2.6 and Appendix Figure D.1d). In sample P2-TN1, the only predicted driver point mutation was a deleterious mutation in the tumour suppressor gene *TP53*, and this was detected in both DCIS and invasive compartments and was also present in all cancer regions of the second PBC, P2-TN2, consistent with an early onset in the development of this cancer (phylogenetic tree; Figures 4.7 and 4.8). These data therefore suggest that many, if not all, of the genetic events necessary to initiate the invasive transition in these three cancers were acquired within the ducts, and subsequently both intraductal expansion and stromal invasion ensued.

4.2.4. Phenotypic changes accompany progression

Next, by integrating additional layers of spatial data, we sought to establish how phenotypic changes relate to genetic-state and histological-state transitions. In P1-ER1 and P1-ER2, consistent with a more proliferative phenotype, *PTEN*-mutant clone regions exhibited denser Ki-67 IHC nuclear staining, than *PTEN* wild-type ancestral clone regions (FDR = 0.004 P1-red versus P1-orange; and FDR = 0.03 P1-purple versus P1-green) (bottom panel Figure 4.7 and Appendix Figure D.1e). However, for a given genetic clone, the Ki-67 score was similar irrespective of whether it occupied a DCIS or invasive state, indicating that upregulation of Ki-67 is temporally related to acquisition of a *PTEN* mutation and precedes invasion.

By contrast, cellular resolution spatial transcriptomics analysis of P1-ER2 revealed that epithelial cell expression of several genes – *CLDN4* (encoding claudin 4), *ACTB* (encoding β-actin), *KRT5* (encoding keratin 5) and *CTSL2* (encoding lysosomal cysteine protease cathepsin V) – differed between DCIS and invasive compartments occupied by the same, P1-purple, clone (Appendix Figure D.1f). These transcriptional changes might therefore be considered more closely linked to the histological transition rather than genetic changes traced by this approach. Expression of *CLDN4* was consistently lower in the invasive compartment than to each DCIS clone. However, for some genes such as *ACTB*, expression patterns changed in opposing directions in the invasive cancer relative to the sampled DCIS clone (expression is higher than P1-green DCIS (FDR = 0.02) and lower

than P1-purple DCIS (FDR = 0.013)) or were highly specific to a genetically more distant DCIS clone (Appendix Figure D.1f).

Attempts to isolate the changes associated with invasive transition might also be confounded by heterogeneity within the invasive compartment. In P2-TN1, we therefore sought to examine whether the two genetically distinct invasive subclones (P2-blue and P2-purple) were phenotypically distinct. The two cancer clones exhibited distinct morphological (nuclear and architectural) features (P = 0.04, Fisher's exact test) (H&E image insets; Figure 4.8) and occupied neighbourhoods with different stroma (FDR = 0.02) and immune cells such as myeloid cell densities (FDR = 0.08) (mini-image insets; top panel Figure 4.8 and Appendix Figure D.2a–c). Transcriptional programs were also distinct, with statistically significant differences in gene expression for 12 of 91 genes between clones (Appendix Figure D.2d). Together, these data indicate that the particular clones sampled can have a profound effect on attempts to identify the phenotypic changes implicated in driving or arising during histological progression.

4.2.5. Growth patterns of pre-invasive clones

To demonstrate that BaSISS can be used to chart growth patterns in relation to complex tissue structures, we turned our attention to three DCIS samples from P1 that spanned a tissue surface area of 224 mm² (P1-D1, P1-D2 and P1-D3) (top panel Figure 4.9 and Appendix Figure D.3a). The adult female breast comprises multiple, branching ductal systems, termed lobes, that extend from the nipple surface to the acini of the lobules [Going et al. 2004; Schnitt et al. 2013]. DCIS arises from the duct epithelium and is considered a lobar disease as it typically involves the ducts and lobules of a single lobe [Pinder 2010]. Although DCIS is known to be genetically heterogeneous [Casasent et al. 2018], how DCIS clones are organised and grow through the wider duct system remains elusive [Thomson et al. 2001].

The clone maps generated for the three samples formed striking mosaics of mainly green and orange, and occasional blue and grey that localized to areas of histologically confirmed DCIS (top panel Figure 4.9 and Appendix Figure D.3a). Immune clusters and occasional normal or hyperplastic ducts appeared white (unstained), consistent with a different genetic ancestry. In P1-D3, a 3-mm length of a large duct exhibited both a genetic and a histological transition from normal ductal epithelium to DCIS along its length, confirming that, although neoplastic involvement was extensive in this lobe, it was incomplete (Appendix Figure D.3a). On dividing the glandular tissue into lobules (white dashed contours; Figure 4.9),

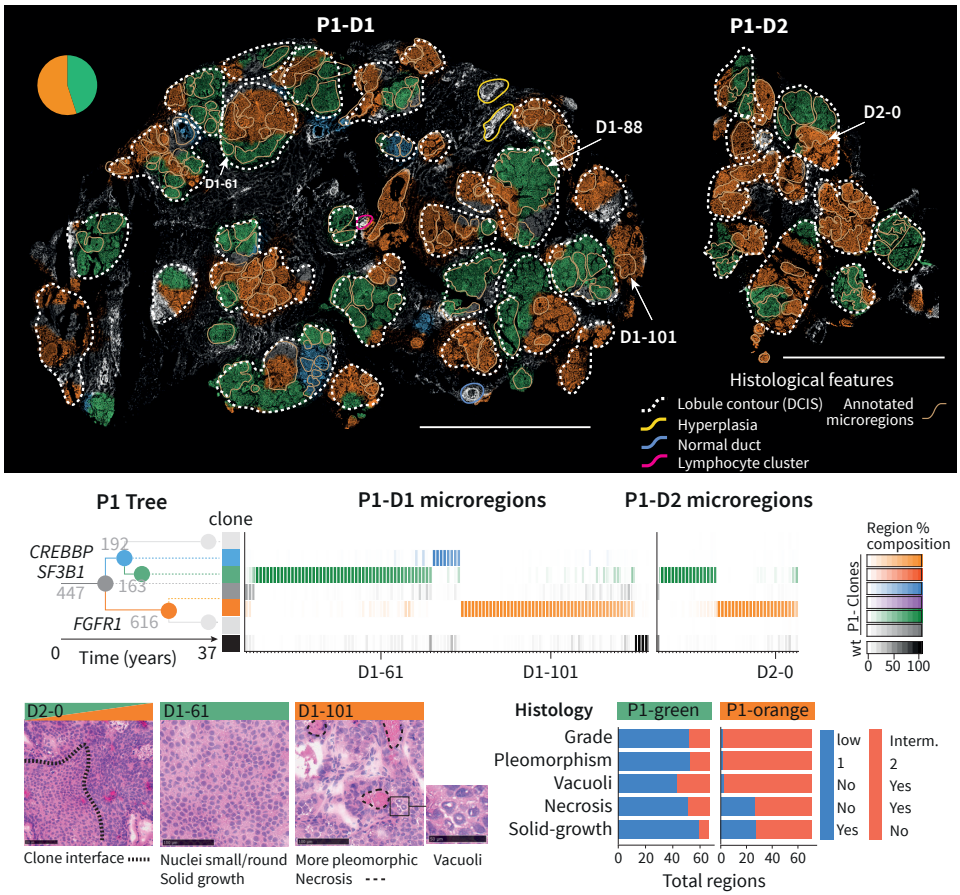
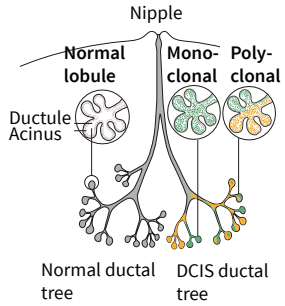


Figure 4.9: Growth patterns and histological associations of DCIS clones. ***top*** BaSISS maps of pure DCIS samples: P1-D1 and P1-D2. The most prevalent genetic clone is projected as a coloured field (which corresponds to ***middle***) on DAPI images (reported if the CCF > 25% and the inferred local cell density > 300 cells per mm^2). Scale bar, 5 mm. The pie chart reports the WGS-estimated clone composition. The white dashed contours delineate morphologically defined lobules. The beige contours mark 114 and 40 manually selected microregions in P1-D1 and P1-D2, respectively, the clonal composition of which is reported by the heatmaps in ***middle***. Microregions were manually selected and represent single or small groups of intimately related acini or ductules from the same lobule. ***middle***, The phylogenetic tree inferred from P1 multi-region WGS. Branches and nodes are coloured to reflect the clone maps. Only branches detected in P1-D1 and P1-D2 are coloured. WT - wild type. ***bottom-left***, H&E images report representative subclone histological features in regions selected from ***top***. Scale bars, 100 μm and 50 μm (vacuoles). ***bottom-right***, Stacked bar plot summarizes histological features of microregions dominated by P1-green (n = 66) or P1-orange (n = 72). Nuclear pleomorphism is a measurement of the amount of variability in size and shape of the nuclei and is a major determinant of the histological grade.

it was apparent that a handful of lobules contained a single clone, but often multiple clones co-occurred. Indeed, we were surprised to observe that the same clones repeatedly co-existed within lobules that spanned centimetres of tissue. These appearances seem at odds with the traditional model of clonal competition in which a fitter clone generates localised monoclonal sweeps.



Cartoon of a lobe of the breast with normal anatomy (left) and DCIS (right), with lobules exhibiting monoclonal and polyclonal involvement.

However, at finer, sublobular resolution, complete or near-complete clonal sweeps are the dominant pattern, as exemplified by assaying 146 representative microscopic regions that represent individual or small clusters of intimately related acini and ducts (beige contours; Figure 4.9 top). The existence of frequent clonal sweeps as inferred by BaSISS (Figure 4.9 middle) was corroborated by LCM–WGS of additional microregions (Appendix Figure D.3b). In some instances, including P1-D1-88 (Appendix Figure D.3c) and P1-D2-0 (Figure 4.9 and Appendix Figure D.3d–f), clonal interfaces are directly observed within a continuous anatomical space. However, more commonly, rapid clone field transitions (see interactive maps at https://www.cancerclonemaps.org/) coincided with the myoepithelial cell layer and/or basement membrane that define an acinus or ductule border. It thus transpires that the microanatomical structure of resident tissues can have, an as yet poorly understood, role in shaping observed subclonal architectures (top panel Figure 4.9).

4.2.6. DCIS clone-specific phenotypes

Integration of histological and spatial gene expression data from serial sections revealed that the DCIS clones, P1-green and P1-orange, exhibit many phenotypic differences that are consistent across large tissue areas (Figure 4.9 and Appendix Figures D.3 and D.4). Histogenetic associations were very strong, with regions dominated by P1-green being more likely to have an intermediate rather than a low nuclear grade ($P<0.0001$; Fisher's exact test after Bonferroni correction), exhibit more nuclear pleomorphism ($P < 0.0001$), necrosis ($P < 0.0001$), vacuoles ($P < 0.0001$) and a non-solid architectural growth pattern ($P < 0.0001$) (Figure 4.9 and Appendix Figure D.3).

Clone and cell type-resolved spatial gene expression analysis using targeted ISS further corroborated phenotype–genotype correlations. A total of 28 of 91 interrogated genes were differentially expressed by the two main clones (FDR < 0.1, fold change > 1.5 both ways; Appendix Figure D.4). Consistent with a higher nuclear grade, P1-orange epithelial cells exhibited higher expression of the cell-cycle regulatory oncogenes *CCND1* and *CCNB1* and the oncogene *ZNF703*, which have been linked to adverse clinical outcome [Solin et al. 2013]. Overall, architectural

and nuclear appearances and gene expression profiles were remarkably lineage-specific, and it was particularly notable that these different patterns could also be appreciated spatially, in regions with sublobular, microscopic clone intermixing, adding weight to the clone composition predictions by the model (Appendix Figure D.3d).

4.2.7. Metastatic clones in a lymph node

Lymph node metastasis is associated with higher rates of cancer mortality [Jatoi et al. 1999]. Whether it has an active role in facilitating cancer progression or simply reflects a more aggressive or distinct biology of certain clones is largely unknown. A substantial challenge is low cancer purity of diffusely infiltrated lymph nodes, which can make it difficult to separate cancer from immune cell-derived molecular signals. To demonstrate that BaSISS can facilitate the simultaneous study of cancer and immune compartments in such challenging cases, we analysed BaSISS, histological annotation and ISS targeted gene expression datasets from sample P2-LN1 (Figure 4.11 and Appendix Figures D.5 and D.6).

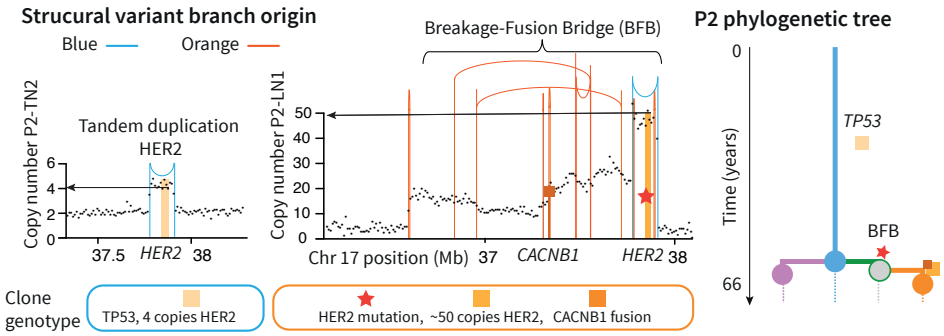


Figure 4.10: genomic structures in P2-blue and P2-orange clones. Plots of the genomic structures in P2-blue and P2-orange clones in the vicinity of the *HER2* gene, derived from WGS data of P2-TN2 and P2-LN1. Vertical lines represent genomic rearrangement breakpoints coloured by the phylogenetic tree branch where the event occurred. Dots represent local (binned) copy number. *HER2* amplification, *CACNB1* fusion and *HER2* mutation are BaSISS targets used to track this complex event. BFB, breakage-fusion-bridge.

BaSISS in P2-LN1 targeted 13 trunk and branch alleles, including point mutations and an expressed novel internal fusion in the *CACNB1* gene that was co-amplified with the clinically targetable breast cancer oncogene *HER2* in a breakage-fusion-bridge (BFB) event (Figure 4.10 and Supplementary Data Table 1). The model detected two clones (P2-blue and P2-orange) that formed

spatially segregated patterns in P2-LN1 (Figure 4.11). Only P2-blue was detected in primary breast tumours (P2-TN1 and P2-TN2) (Figure 4.8 and Appendix Figure D.2b).

Detailed histological annotation, blinded to the clone territories, was performed using a combination of H&E, CD45 and pan-cytokeratin IHC and identified multiple metastatic cancer growth patterns (coloured contours; Figure 4.11). Intersecting the clone maps and histological annotations revealed strong associations between the two detected clones and the two main histological growth patterns ($P < 0.0001$, Fisher's exact test) (Figure 4.11 top-right). The P2-orange clone formed monotonous sheets of cancer cells, exhibited weak immunoreactivity for pan-cytokeratin and often occupied sinusoidal structures. By contrast, P2-blue cells stained more strongly for pan-cytokeratin and, when clustered, surround densely packed lymphocyte cores (Figure 4.11 right and Appendix Figure D.5a–d).

We sought to determine whether transcriptional differences support the spatial inference of clones. Consistent with the known HER2 amplification, P2-orange expressed higher levels of HER2 (bottom panel Figure 4.11 and Appendix Figure D.5c). A total of 17 of 91 genes were differentially expressed and many of these are implicated in critical biological cancer pathways and/or have recognized prognostic value, including *CTSL2*, *VEGFA* (encoding vascular endothelial growth factor receptor A) and CD24 [Kwon et al. 2015; Sereesongsaeng et al. 2020] (bottom panel Figure 4.11). Spatially plotting these genes confirmed that clone-specific expression patterns are recapitulated within multiple, spatially distinct expansions across more than 1 cm^2 of tissue (Appendix Figure D.5a–c).

Integration of spatial transcriptomics data also revealed that metastatic subclones occupied distinct immune microenvironments. Relative to P2-orange cells, P2-blue cells resided in neighbourhoods enriched for T cells and B cells (bottom panel Figure 4.11). In fact, P2-blue cells frequently formed clusters around B cell-rich germinal-like centres, highlighting a potential clone-specific interaction with the adaptive immune system (middle-right panel Figure 4.11 and Appendix Figure D.5a,d). By contrast, P2-orange regions frequently resided inside the lymph node sinuses that were lined by endothelial cells expressing *CD34* and *PDGFRB* (middle-right panel Figure 4.11 and Appendix Figure D.6). Most of the immune cells in P2-orange regions were myeloid cells with expression profiles consistent with the presence of both M1 and M2 macrophages (*CD163*, *CD68*, *HAVCR2* and *FCGR3A*), and the most highly enriched gene, *CXCL8*, is released by hypoxic macrophages [X.-P. Li et al. 2015] (bottom panel Figure 4.11). Indeed, relative to P2-blue, it emerges that P2-orange experienced more hypoxic conditions mani-

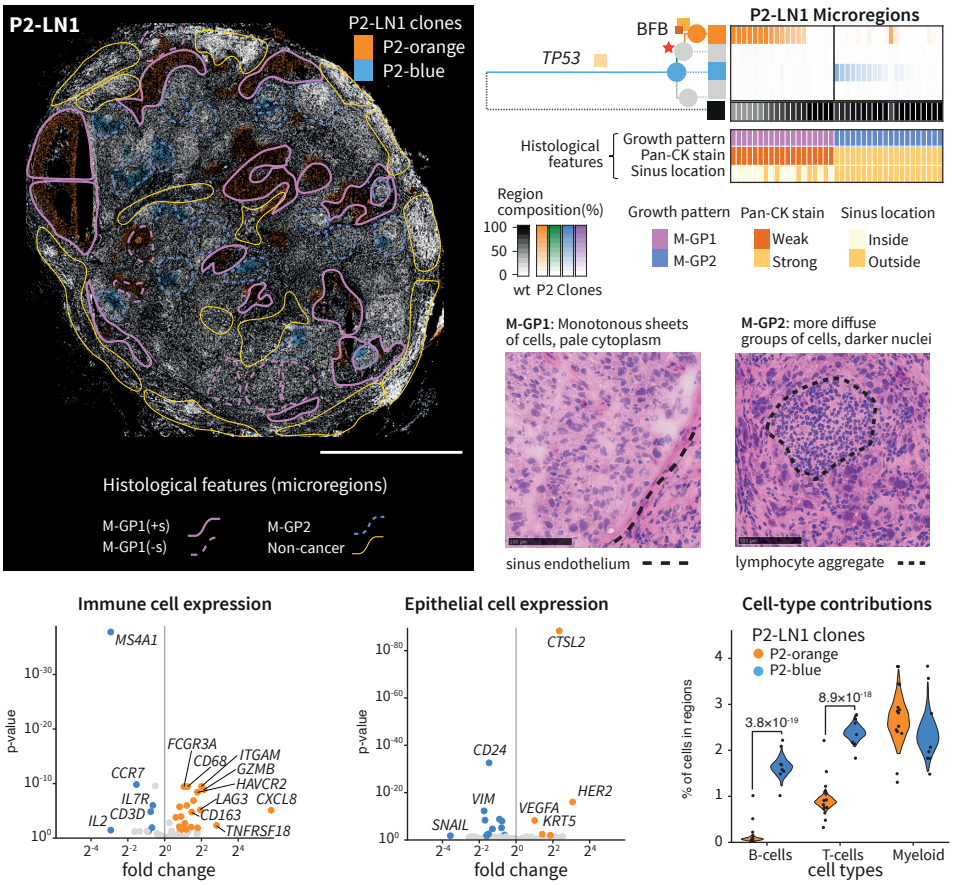


Figure 4.11: Intrinsic and extrinsic features of metastatic subclones in a lymph node. ***top-left***, BaSISS map of P2-LN1, which relates to P2-TN1 (Figure 4.8) and P2-TN2 (crefappfig:appendix-applications-PBC-P2a,b). The most prevalent genetic clone colours are projected as coloured fields on the DAPI image (reported if the CCF > 25%; a threshold of 5% is used in regions of diffusely infiltrating blue to allow visualization in very high normal contamination regions). Scale bar, 2.5 mm. Coloured contours define microregions with distinct metastatic cancer growth patterns (M-GP1 and M-GP2); '+' indicates the surrounding sinus epithelium. ***top-right***, Phylogenetic tree inferred from P2 multiregion WGS. Branch and node colours inform the clones mapped in ***top-left***. *HER2* amplification (orange squares: pale - low, bright - high amplifications), *CACNB1* fusion (red square) and *HER2* mutation (red star) are BaSISS targets used to track this complex event. The top heatmap reports the BaSISS clone contribution to 39 histologically annotated microregions from a (regions with 5% or more tumour cells are included). The bottom heatmap reports microregion histological features. Pan-CK, pan-cytokeratin. ***middle-right***, Representative areas of the two main growth patterns stained with H&E. Scale bar, 100 μm. ***bottom***, From left to right: volcano plot of immune cell expression of the 62 genes in the ISS immune panel, volcano plot of epithelial cell expression of the 91 genes in the ISS immune panel. Significantly (FDR > 0.1), differentially expressed (fold change of more than 1.5 both ways) genes are coloured. Violin plots depict clone-specific cell-type contribution posterior density of the generalized linear mixed model with region-specific random effect, and includes the 22 clone territories with a dominant clone fraction > 0.05 in P2-LN1. Significant comparisons were controlled for FDR using the Benjamini–Hochberg procedure.

festing as higher cancer cell expression of *VEGFA* and necrotic regions (Appendix Figure D.6). Hypoxia signatures are associated with adverse clinical outcomes, probably because they reflect the emergence of environments that can select for hypoxia-tolerant clones and/or cancer proliferation rates outstrip neoangiogenesis [R. A. Cairns et al. 2004]. Together, these data demonstrate how BaSISS clone maps allow one to spatially relate such variation in microenvironments to individual clones.

4.3. Discussion

Applying BaSISS to a series of samples from the key stages of breast cancer progression – carcinoma *in situ*, invasive cancer and lymph node metastasis – it is notable that virtually every sample exhibited a spatial organisation of clones, which warrants further investigation in larger cohorts. The fact that nearly all clones examined in this dataset displayed distinct clone-specific gene expression, stromal and immune microenvironments and microanatomical niches highlights the functional relevance of at least some subclonal diversification.

A key observation from these data is the variable relationship between genotype and phenotype. We examined three primary breast tumours with intermixed DCIS and invasive disease, and in each case, a clone was identified that existed simultaneously in the two distinct histological states. In contrast, albeit in a single examined case, we observe remarkable consistency between genetic lineage and DCIS grade spanning centimetres of tissue. While occasional examples of 'upgrading' occur, which could be accounted for by unsampled genetic evolution, we find no evidence of transition to a lower grade. The lack of plasticity within the grade phenotype space is aligned with studies that report that DCIS and invasive cancers are usually of the same grade [Gupta et al. 1997; Luijt et al. 2016]. Our findings could reflect stable differentiation states acquired by divergent cells, prior to DCIS onset [Rakha et al. 2022].

Another key message, that echoes those of a recent spatial study in melanoma [Nirmal et al. 2022], is that measuring entire tissue sections rather than small fields of view is advantageous for understanding biology. Firstly, this approach reveals that patterns of cancer growth and organisation can vary across different scales. For example, in P1-D1, P1-D2 and P1-D3, at macroscopic scale we see DCIS clones co-exist in multiple lobules, across an entire breast lobe, while at the microscopic level they are mainly segregated, distending individual acini or forming abutting populations. There are various possible explanations for the observed

DCIS growth patterns such as clonal co-operation or genetic drift [Janiszewska et al. 2019; Turajlic et al. 2019]. A number of recent mathematical modelling studies have emphasised the importance of tissue architecture in shaping evolution and in the reported case it is certainly plausible that altered rates of fixation and absorption arising from the physical bottlenecks in the different sized ducts and ductules could account for the observed genetic patterns [Lieberman et al. 2005; Noble et al. 2022; West et al. 2021]. Secondly, within a lymph node the panoramic view allows one to appreciate that subclones repeatedly foster the same ecosystems – one that is immune depleted within the sinuses while the other clustered around B-cell aggregates consistent with a clone specific interactions with the adaptive immune response [Sharonov et al. 2020]. This way of examining tissues naturally leads to a multitude of questions that each warrant follow up in its own extended cohort. The tools we have developed and approaches demonstrated in this study will provide others with a framework for achieving this.

The ability to chart clonal growth patterns and clone-specific genetic underpinnings of the tumour microenvironment is likely to be instrumental in elucidating how different evolutionary processes operate and manifest across different cancer types—or even in histologically normal tissues [Sottoriva et al. 2015]. Understanding the forces of malignant progression, especially invasion and metastasis, and how interactions with the tumour microenvironment shape clinical outcomes [Risom et al. 2022] appear of particular importance. Detailing the functional and microenvironmental characteristics of different clones is also relevant as some part of subclonal diversity in tumours may be due to selectively neutral drift, but the exact extent remains debated.

Particular advantages of the technology are that it is capable of interrogating very large tissue sections on the scale of squared centimetres, which enables studying entire cross-sections of smaller tumours. It is also comparably cheap, unlike solely relying on sequencing-based methods [Vickovic et al. 2019]. The three main limitations of the approach are relatively low sensitivity, which currently precludes single-cell genotyping, a reliance on RNA with the resulting variation in gene expression levels of targeted transcripts, and the fact that clone-defining mutations need to be detected first by separate sequencing-based assays. Greater sensitivity and spatial resolution may be achieved by including more targets per clone and by favouring mutations with higher predicted expression levels, for example, in higher copy number states. A switch to hybridization-based sequencing and direct RNA-binding probes may also improve base-specific detection by several fold [Gyllborg et al. 2020; H. Lee et al. 2022].

It is often stated that ‘nothing in biology makes sense except in the light of

evolution' [Dobzhansky 1973], which is likely to be true for cancer biology. The ability to spatially locate and molecularly characterize different cancer subclones adds essential features to the spatial-omics toolkit. It provides a robust evolutionary framework that is necessary to interpret the biological relevance of many of the more plastic spatial characteristics of a cancer. Future widespread applications of spatial genomics approaches such as BaSISS will uncover how cancers grow in different tissues and allow us to track, trace and characterize the ill-fated clones that are responsible for adverse clinical outcomes.

Outlook

5

5.1. Summary

Cancer is fundamentally a spatial disease as its origin and progression are linked to specific anatomical locations and microenvironments within the body. These spatial relationships between cancer cells and adjacent tissues offer key insights into tumour onset, evolution, and treatment response (Chapter 1). Despite this importance, current technologies constrain our understanding of the interactions between subclones and the microenvironment. Existing spatial technologies targeting the genome either suffer from limited spatial resolution or employ indirect methods based on a limited subset of genomic changes, thereby complicating phylogenetic inference.

In this thesis, I developed the key computational components for one of the first method to chart cancer evolution in tissue space (Chapter 2). I designed algorithms that integrate clonal maps with various types of auxiliary data, including ISS, IHC, and histology, to provide a comprehensive characterisation of clones and the tumour microenvironment (Chapter 3)

Applying these methodologies to multifocal breast cancers, I charted the spatial organisation of clones through three stages of cancer progression: ductal carcinoma *in situ* (DCIS), invasive carcinoma, and lymph node metastasis (Chapter 4). This study revealed a remarkable degree of segregation among histologically distinct cancer clones within the breast's ductal system. It tracked the phenotypical changes accompanying progression from CIS to the invasive stage and confirmed that genetically indistinguishable clones can exist in both invasive and *in situ* states. The study also showed how cancer clones adapt to unique ecological niches within the lymph node, suggesting possible interactions with the immune system. Across all three stages examined, the territories occupied by different clones exhibited distinct transcriptional and histological

features, as well as diverse microenvironments.

These results demonstrate the utility of spatial genomics to elucidate the mechanisms that govern cancer evolution and spatial ecology.

5.2. Current limitations and future solutions

The approach comprises two distinct components: lineage tracing with BaSISS and the characterisation of cell phenotype and microenvironment. As these components operate independently and possess their own specific limitations and potential solutions, I will discuss them separately.

5.2.1. Spatial lineage tracing with BaSISS

While the BaSISS protocol offers the unique ability to map cancer clones across square centimetres of tissue, there are several limitations that are important to consider. Some of these are intrinsic to the technology, such as its emphasis on mutations in the coding region and its dependency on prior mutation detection and phylogenetic reconstruction. Other limitations, such as the lack of detailed phylogeny, low signal yield, and standardisation issues, could possibly be addressed with existing technology. Overcoming these constraints would improve both the resolution and the granularity of cancer evolution characterisation.

Prior knowledge of genome alterations

BaSISS is a targeted method that require both prior knowledge of the mutations and requires a panel redesign for every individual cancer case, drastically increasing the cost of new experiments. This contrasts with spatial sequencing methods [Erickson et al. 2022; Zhao et al. 2022] which do not require prior information and infer clone maps based on copy number profiles. While it is possible to adapt BaSISS to this strategy by designing probes for a set of expressed genes distributed across the genome, such an approach would only yield copy number profiles. Consequently, this would forfeit BaSISS's primary advantage: the ability to detect point mutations.

Prior phylogeny reconstruction

Initial detection of subclones as well as tree reconstruction was done based on multi-regional sequencing data. This method is often not sensitive enough to resolve clones in cases where the CCF is low. An alternative approach to improve

point mutation detection and tree resolution could involve techniques like single-cell WGS. Currently, LCM might offer an even better option, owing to its superior ability to detect point mutations.

When employing LCM one could slightly modify the research workflow. Instead of first inferring cancer clones and subsequently exploring their association with histology, histological features could guide the selection of LCM regions for genotyping. While this strategy introduces some bias, it could more effectively elucidate the relationship between histology and genomics. BaSISS could then extend genotyping capabilities to regions not characterised by LCM. For an illustration of the additional phylogenetic details that can be revealed through such methods, refer to Figure 4.3.

Subclone resolution

The number of probes defining clonal expansion is critical for accurate subclone mapping. For instance, the P1-green clone (Figure 4.9) was defined solely by a point mutation in the *KIAA0652* gene. Fortunately, this gene was highly expressed, mitigating the risk of failing to resolve this particular clone. Although gene expression levels were taken into account during probe design, predicting gene expression patterns is challenging. This is particularly true when expression varies non-linearly with changes in copy number (see Section 2.4 for details). Additionally, it is difficult to forecast the efficacy of padlock probes, such as their specificity. Therefore, it is crucial to include as many genomic alterations as possible in the probe design.

A fundamental limitation of the BaSISS approach, and any other method targeting RNA, is the inability to trace clones not defined by mutations in the coding region. The mutation burden can vary considerably across different cancer cases and types, with the number of coding mutations per tumour ranging from 1 to 10,000 [Martincorena and Campbell 2015]. As a result, some tumours are more compatible with BaSISS analysis. In particular, hypermutant tumours, which contain a significantly higher number of mutations compared to typical cases, may be especially amenable to investigation, given the greater likelihood that their clones will accumulate mutations in the coding region.

When the task involves tracing hundreds of mutations, accomplishing this with BaSISS is relatively straightforward. The 56 probes that were used to map the P1 clone are not a technological limitation for the ISS protocol. Modern ISS experiments can resolve up to 300 genes, which, in the context of BaSISS, equates to the detection of 150 allelic variants plus an additional 150 reference alleles.

Signal yield

A significant drawback of the BaSISS protocol is its low yield of signals, which arises from two main factors: inefficient sequencing-by-ligation chemistry and limited targetable nucleic acids. Unlike standard FISH technologies, where probes can target multiple regions of the same gene to enhance detection, BaSISS can target SNVs and structural variants with exactly one probe.

However, there is room for improving the sequencing chemistry, as demonstrated in newer versions of ISS protocols. For instance, hybridisation-based ISS (HybISS) [Gyllborg et al. 2020] and direct RNA-targeted ISS (dRNA-HybISS) [H. Lee et al. 2022] offer improvements. The former one exchanges sequencing by ligation with a more efficient hybridisation step. The latter bypasses the need for reverse transcription required for cDNA generation. Nevertheless, it is essential to consider that RNA ligases used in these improved methods may have different substrate specificities, which could impact ligation accuracy and the off-target effect.

Standardisation

Standardisation of the BaSISS protocol could substantially enhance its utility and reproducibility across different laboratories. The BaSISS protocol involves a diverse array of equipment, reagents, and computational infrastructure, all of which must be meticulously operated. Additionally, trained personnel are required to execute each stage with minimal variability. Even within the scope of this thesis where the experiments were conducted by the same individuals, I observed considerable variability in signal detection between slides. Attempts to replicate the experiments in a different facility have highlighted the challenges associated with even minor issues, such as imperfect alignment of the glass slide under the microscope between reaction cycles. These seemingly trivial errors took months to identify and led to significant degradation in data quality. Transitioning the protocol to industrial-standard machines, similar to those used in the 10x Xenium technology, could make the method more reproducible and accessible to a broader research community.

Scope of future applications

Is this technology a silver bullet for solving the challenges of spatial genomics? Probably not yet. Despite its many advantages, the BaSISS approach has some inherent drawbacks. While it is cost-effective for profiling multiple samples from

the same tumour, the need for customised panel design for each experiment to capture SNVs elevates the costs. It improves upon RNA sequencing methods which can only resolve clones based on copy number profiles; however, BaSISS still loses to LCM-WGS due to its fundamental constraint: the necessity for clone-specific mutations in the coding region, which are limited. Additionally, despite the avenues for improvement, this approach currently suffers from low sensitivity and lacks single-cell level genomic characterisation. Consequently, spatial lineage tracing remains a significant challenge.

Despite the limitations, profiling the clonal structure of normal tissues remains feasible. Even though nomral cells generally exhibit a lower mutation burden, LCM studies identify up to 100 mutations in their coding regions. Although, one need to keep in mind the age of donors, for example in R. Li et al. [2021] they were 85 to 93 years old.

However, I envisage that BaSISS could serve as a valuable supplement to other spatial genomics tools like LCM (Box 1), particularly for deeply investigating the evolutionary history of hypermutant cancers.

The model I developed (detailed in Chapter 2) along with the accompanying infrastructure code, is designed to handle moderately noisy BaSISS data in future experiments. This model is capable of mapping an arbitrary number of clones across centimetres of tissue, assuming each clone possesses at least one well-behaved, clone-specific mutation. While there is room for improvement, as discussed in Section 2.4, further ameliorations will be needed as experimental methods advance to offer single-cell-level resolution. Such advancements will invite substantial changes to the model's structure. Although the data generation process will share many similarities with the current approach, future data sets will likely not require binning or GP to manage sparsity. I anticipate that future models will employ clustering approaches similar to Bayesor [Petukhov et al. 2022] or pciSeq [Qian et al. 2020].

5.2.2. Phenotype and microenvironment characterisation

The auxiliary data acquired for the study presented in this thesis is largely suboptimal by current standards. This limitation arises from the fact that the experimental setup was originally designed in 2016[1] by L.R.Y., J.S., M.N., and P.J.C., predating the rapid advancements in the field of spatial omics. Consequently, the workflow detailed in Chapter 3 should be viewed primarily as a proof of concept and should be replaced with modern methodologies in future research.

[1] It is strange to think that the original experiments took place just as I began my undergraduate studies.

Histological characterisation

Histological features offer a straightforward method of assessment, requiring often only simple staining to reveal a wealth of information. A trained histopathologist can interpret cell arrangements and nuclear shapes, which often suffice to define the micro-anatomy of the tissue that impose structural constraints on the tumour's environment (Section 1.2). Despite its utility, histopathological annotation performed by human is labour-intensive, subjective and difficult to scale.

Digital histopathology offers an alternative approach. Besides feature extraction, it is possible to conduct semantic segmentation of the regions and extract its micro-anatomical structures [Kiemen et al. 2020]. Recent studies employing artificial intelligence have demonstrated that the histological characteristics of tumour cells correlate with underlying genomic alterations [Fu et al. 2020; Kather et al. 2020]. However, a significant challenge lies in integrating slide-level bulk genomic data with tile-based digital features [Shmatko et al. 2022]. Spatial genomics can address this issue, provided that a sufficiently large dataset is generated.

Spatial proteomics and transcriptomics

Spatial transcriptomic and proteomic technologies have undergone significant advancements in recent years [Lewis et al. 2021; Mund, Brunner et al. 2022]. This progress has made the acquisition of single-cell level proteomics and transcriptomics increasingly routine (Section 1.3 and Box 1). Although these technologies focus on different modalities – for example, protein profiles more readily identify immune cells – they yield rich data sets. These data are sufficient for mapping the spatial distribution of cell types at, or slightly above, the single-cell level.

Highly multiplexed antibody detection methods yield rich data sets that facilitate the identification of functionally recurrent neighbourhoods within cancer tissue [Danenberg et al. 2022; Jackson et al. 2020; Nirmal et al. 2022; Schürch et al. 2020; X. Q. Wang et al. 2023]. Given that proteins serve as primary mediators of cell-cell interactions, these technologies also enable the direct observation of cell-cell contacts, providing unequivocal evidence of interactions [Nirmal et al. 2022; X. Q. Wang et al. 2023].

To date, spatial transcriptomics assays for cancer have primarily employed 10x Visium, a technology that enables whole-transcriptome profiling and the mapping of various niche cell types, guided by rich single-cell transcriptomics reference data [Andersson et al. 2021; Barkley et al. 2022; Berglund et al. 2018; Erickson et al. 2022; Gouin et al. 2021; Ji et al. 2020; Moncada et al. 2020; Qi et al. 2022;

Ravi et al. 2022; Ståhl et al. 2016; R. Wu et al. 2021; S. Z. Wu et al. 2021]. However, because these methods operate at a supercellular level, they are mostly limited to describe broad patterns of co-occurrence and offer only indirect evidence of functional interactions (Figure 1.5). The recent commercialisation of fluorescent probe-based technologies, such as 10x Xenium and MERSCOPE, is likely to shift this landscape, allowing for transcriptional characterisation at the single-cell level.

Overall, the success of spatial transcriptomics and proteomics, relative to other spatial modalities, will likely shape the field's trajectory towards phenotypic characterisation of cell communities in the coming years.

Towards comprehensive spatial models of carcinogenesis

The statistical models outlined in (Section 3.2.6) serve to delineate compositional and expression differences between preselected regions occupied by distinct clones. With the availability of richer single-cell level features, compositional differences can be clustered into recurrent neighbourhoods using methods such as topic modelling or graph-based models, further enriching the description of tumour microenvironment [Danenberg et al. 2022; Jackson et al. 2020; Nirmal et al. 2022; Schürch et al. 2020; X. Q. Wang et al. 2023].

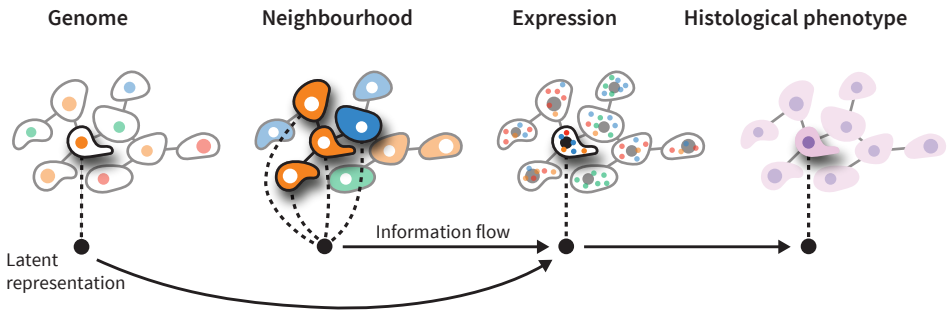


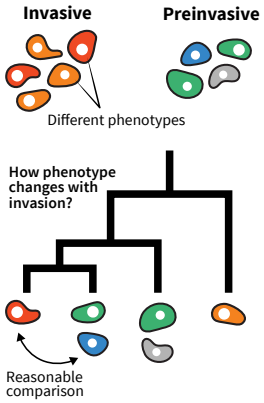
Figure 5.1: Conceptual comprehensive spatial model. Multiple spatial modalities at the single-cell level offer rich data for analysing cancer cell biology. A graph-based framework provides an effective approach for integrating these data types. Inspired by Fischer et al. [2023]

However, richer multimodal single-cell data enable more refined statistical approaches for modelling interactions between intrinsic and extrinsic cellular components. Graph-based representations preserve spatial relationships and offer a framework for modelling intercellular communication, as demonstrated by the NCEM method [Fischer et al. 2023]. While fully mechanistic modelling of cellular

components poses significant challenges, neural network-based approximations could partly represent cell behaviour. I anticipate that the integration of multimodal data within a graph-based model, which explicitly considers cell phenotypes as functions of both their intrinsic characteristics and local neighbourhood, will advance our understanding of the roles of genome and microenvironment in cancer cell behaviour and cellular plasticity (Figure 5.1).

5.3. Opportunities for studying breast cancer

The study detailed in Chapter 4 demonstrates a surprising degree of spatial organisation among cancer clones. This work highlights the efficacy of cancer lineage tracing technology in establishing links between evolutionary lineages, their phenotypes, and their microenvironments. This technology thus offers valuable insights into cancer cell plasticity from an evolutionary perspective. While most of the findings are discussed in Section 4.3, this section aims to situate these results within the broader context of unresolved questions concerning breast cancer progression, closer to the one listed in Section 4.1.6.



Understanding the evolutionary underpinnings of phenotypic variability aids in making accurate comparisons to elucidate biological questions like "What phenotypic changes accompany the invasion process?"

Mammary gland development and ductal system colonisation

Understanding early evolution and clonal dynamics within the ductal system is crucial for grasping how breast cancer develops its malignant potential. Previous studies have reported extensive heterogeneity of carcinoma *in situ*, however they did not address their precise spatial distribution over the large tissue areas [Casasent et al. 2018; Nishimura et al. 2023; Yates et al. 2015].

The observed patchwork pattern of microscopic clonal sweeps within the DCIS is striking and prompts questions about its formation. Several explanations for these growth patterns exist, including clonal co-operation and genetic drift, as discussed in Section 4.3 [Janiszewska et al. 2019; Turajlic et al. 2019].

This observed pattern also aligns with simulations conducted by West et al. [2021], which predict that local clonal sweeps arise due to the spatial constraints inherent in the fine branches of the ductal system. However, this model considers the mammary gland as a static system, neglecting its dynamic nature and life-long remodelling (Section 4.1.1). Developmental processes within the mammary gland could play a role in the clone spread.

The first detected clonal expansion in this case roughly coincides with puberty, a period of active ductal system formation. This expansion could have been facilitated by such developmental processes, stimulating cancer cells to proliferate and

disseminate throughout the newly formed ductal and lobular units. Subsequent expansions may also align with other periods of breast remodelling, such as pregnancy, although we lack data for the P1 case in this context.

Within the DCIS, we observed a distinct frontline between cancer clones without many examples of intermingling. This observation is intriguing from a clonal dynamics standpoint as it raises questions as to what extend a 'fitter' clone may outcompete others colonising ductal system and why do we observe complete local sweeps. The human breast continues to develop acini with each menstrual cycle [Javed et al. 2013], providing additional space for proximal clones to expand. Once colonised, lobules may become inaccessible to other clones, a phenomenon observed in bacterial colonisation of porous landscapes [Conwill et al. 2022].

Determining the extent to which this pattern results from cancer spreading within a static system versus the role of natural breast development, is challenging based on a static 2D snapshot of a dynamic 3D structure. However, the BaSISS approach can be applied to consecutively capture the three-dimensional structure of clonal distribution. Such data should clarify the degree of clonal mixing at a global level, thereby providing a foundation for future evolutionary simulations. Additional genomic studies of individual lobules, e.g. with LCM may help dissect clonal dynamic further and understand local patterns of colonisation.

Cancer plasticity and invasion

Prior research suggests that genetically identical clones can coexist in both carcinoma *in situ* and invasive states [Casasent et al. 2018]. Moreover, the microenvironment plays a role in initiating invasion [Risom et al. 2022; Sinha et al. 2021] (Figure 4.4). Nonetheless, it is evident that not all clones possess the capacity for invasion, and specific genetic backgrounds may be required for this capability.

The invasive growth patterns discussed in Chapter 4 support this model. Invasive clones occupy the ductal system, occasionally co-localising with non-invasive clones (Figure 4.7). While the absence of direct experimental evidence makes it challenging to assert whether non-invasive clones lack invasive potential, statistical analysis across large cohorts could reveal correlations between the genetic backgrounds of invasive and non-invasive lineages. Such analysis could also quantify associated changes in cell phenotype.

It is plausible that CIS clones correlate with specific microenvironment types, as their histological features differ significantly. Unfortunately, the cell-typing analysis was not sufficiently sensitive to detail the functional differences between

the niches associated with invasive and non-invasive clones, such as those proposed for neutrophils [Sinha et al. 2021] and desmoplastic stroma [Risom et al. 2022]. Future studies, leveraging advancements in spatial transcriptomics and proteomics profiling, should elucidate these microenvironmental differences.

Lymph node metastasis and environment interactions

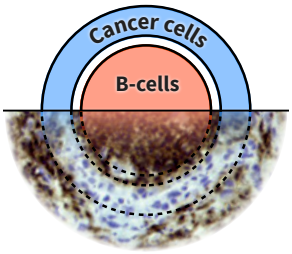
While lymph nodes actively execute adaptive immune responses, their role in cancer progression remains ambiguous in part due to the challenges posed by the low cancer purity in the diffuse infiltration of these structures. Genomic studies have reported extensive heterogeneity among cancer cells within lymph nodes [Bao et al. 2018; Barry et al. 2018; Pal et al. 2021], however they lacked spatial details to functionally link genetic heterogeneity to tumour microenvironment.

The clonal patterns discussed in Section 4.2.7 suggest ongoing evolution within the lymph nodes. However, given that lymph node metastases have been reported to originate from minor clones at primary sites [Bao et al. 2018], these patterns could result from polyseeding by under-sampled clones in the primary tumour.

A striking observation presented in Section 4.2.7 is the specialisation of clones in fostering specific ecosystems. One clone is mostly avoiding immune presence within the sinuses, while another consistently clusters around B-cell aggregates. This indicates clone-specific interactions with the adaptive immune response. To the best of my knowledge, the phenomenon of cancer clones encircling B-cell clusters has not been reported previously. These observations suggest that clone-immune interactions within the lymph node could influence immune cell behaviour at primary sites, either by attracting immune attention or by contributing to immune evasion.

Future spatial studies involving appropriately paired cohorts should further clarify the nature of cancer-immune interactions within both the lymph nodes and primary tumours.

The B-cell Orbital Niche



The P2-blue clone occupies a unique niche around B-cell cluster, as illustrated in Appendix D

5.4. Concluding remarks

This thesis addresses a gap in cancer genomics by introducing a method capable of tracing clones across large tissue sections at single-nucleotide resolution. The computational techniques developed here enable the integration of experimental BaSISS data with multiple phenotypic layers, such as immunohistochemistry, *in situ* sequencing, and histology. Together, these methods enabled the characterisation of individual breast cancer clones and their native microenvironments.

The proposed approach allows for the simultaneous observation of evolving clones, their phenotypes, and the selective environment in which they reside. As such, the utility of this method extends beyond the study of breast cancer progression. At the example of some of CRUK's Cancer Grand Challenges, I will illustrate how this spatial genomics approach could be applied to address other fundamental questions of cancer research. Specifically, the approach could advance the research themes of 'Tumour cell plasticity', 'Normal phenotype', 'Lethal vs non-lethal cancers' and 'Cancer and ageing'.

Cancer cell plasticity. The emergence, spread, and treatment resistance of cancer are often attributed to the ability of cancer cells to alter their states - a phenomenon known as 'cell plasticity'. The origins of this diversity remain unclear. Part of it likely stems from irreversible state changes caused by genetic alterations selected during clonal evolution. Other state changes may be reversible, induced by triggers within the microenvironment. Therefore, methods that enable simultaneous profiling of the genetic composition (intrinsic component), expressed phenotype (a measure of plasticity), and microenvironment (extrinsic component) are necessary to accurately model cell plasticity.

Normal phenotype. Although experiments in model systems suggest that a limited number of driver mutations can initiate malignancy, tissue studies show that cells with multiple driver mutations can still appear histologically normal (Figure 4.3). This points to the significant role of the microenvironment, in conjunction with genetics, in determining tumour phenotype. Notably, cells with identical genotypes can behave differently in varied microenvironments, and vice versa. Therefore, studies that profile genetic composition, microenvironment, and histological appearance are essential for understanding the mechanisms enabling cells to transition between normal and malignant states.

Lethal vs non-lethal cancers. While some cancer clones stay confined within their tissue of origin and thus pose a relatively low risk to the patient, other clones acquire the ability to invade surrounding tissues. A subset of these invasive forms may ultimately metastasise to distant organs, substantially elevating the risk of a fatal outcome. Current screening methods often fail to detect the emergence of these 'lethal' clones, leading to frequent over-diagnosis and subsequent over-treatment. The specific combination of genetic alterations and microenvironmental factors that render a clone 'lethal' remains elusive. Spatial omics techniques hold the potential to not only refine existing histology-based biomarkers but also to discover new biomarkers that could improve patient stratification.

Cancer and ageing. Ageing triggers a cascade of cellular and tissue-level

changes. At the cellular level, it exacerbates genomic instability, leading to a higher mutation rate. At the tissue level, it weakens immune functions and remodels the extracellular matrix. These alterations render tissues increasingly vulnerable to abnormal cell proliferation. However the exact mechanisms of these varied processes remain elusive and spatial genomics and transcriptomics approaches appear well suited to measure how ageing microenvironments affect clonal evolution. Moreover, the processes of ageing may be highly tissue specific, therefore requiring detailed analyses across tissue types to better understand the role of ageing in cancer development.

In conclusion, the spatial genomics techniques developed in this thesis provide a unique opportunity for studying the evolution of cancer clones within their native tissue environment. As demonstrated, utilising these methods to map tumour evolution across various tissues holds great potential for addressing key research challenges in the field of cancer biology.

APPENDIX

Supplementary methods for experimental procedures and data preprocessing

A

Declaration

This supplementary chapter outlines the specific experimental protocols and initial data preparation that generated the data used in this thesis. The content is entirely taken from the supplementary methods section of Lomakin et al. [2022], with small stylistic adjustments.

J.S., P.J.C., M.N. and L.R.Y. designed the initial study and planned experiments. J.S., M.N. and C.S. acquired the ISS data (Appendix A.5). A.S. contributed to image segmentation and processing. A.L.R. and S.S. provided tissue samples (Appendix A.1). S.S. and C.S. performed IHC (Appendix A.2). Young Seok Ju contributed RNA sequencing expertise (Appendix A.3). S.D. performed WGS subclonality analysis (Appendices A.4 and A.7). L.R.Y. conducted LCM-WGS experiments (Appendix A.7). J.S.P, V.V., T.L., Omer Ali Bayraktar and M.Gat. contributed to the development of bespoke ISS analysis pipelines (Appendices A.6 and A.6.3).

While I did not participate in the execution of these protocols — except for the Appendix A.6.2 — this chapter is included for completeness. Detailed experimental protocols and data analysis methods are critical for decisions made in the statistical analysis design, as discussed in Chapters 2 and 3. These details are also important for the interpretation of results presented in Chapter 4.

A.1. Tissue samples

Breast tissue samples were obtained from mastectomies performed for the diagnosis of multifocal primary breast cancer. Patients provided written informed consent for tissue and clinical data use in research studies. Samples and data were obtained and managed in line with the declaration of Helsinki under "project SHARE" #93-085, approved by the Dana-Farber Harvard Cancer Center Institutional Review Board. Sample and data handling at the Wellcome Sanger Institute, Cambridgeshire, UK was performed under the wider framework and approval for the Breast Cancer Genome Analyses for the Inter-

national Cancer Genome Consortium Working Group under REC reference: 09/H0306/36 (Cambridgeshire 3 Research Ethics Committee). The study was later transferred to a protocol REC: 20/PR/0905 approved by London-Harrow Research Ethics Committee.

Tissue blocks were first sliced to obtain sufficient nucleic acids for bulk whole genome sequencing (WGS) and RNAseq analysis. Subsequent serial 10µm sections were obtained from the same tissue block and orientation for base-specific *in situ* sequencing (BaSISS), *in situ* sequencing (ISS), histological assessment and immunohistochemistry (IHC).

A.2. immunohistochemistry

For immunohistochemistry staining after ISS, cover glasses were detached by overnight incubation in TBS-Tween20 (0.05%). Slides were fixed for 10 min with 3.7% formaldehyde (Sigma Aldrich, Munich, Germany). Endogenous peroxidases were blocked with DAKO REAL peroxidase-blocking solution (Agilent, Glostrup, Denmark) for 10 min at room temperature, followed by incubation with DAKO serum-free protein blocking solution (Agilent) for 30 min.

Sections were incubated with mouse monoclonal PanCK antibody (clone AE1/AE3, Agilent) diluted 1:100 in DAKO REAL antibody diluent (Agilent) overnight at 4℃ in a humidity chamber. Thereafter, ready-to-use secondary ImmPRESS HRP Anti-mouse IgG (VectorLaboratories, Orton Southgate Peterborough, UK) reagent was applied for 30min and chromogenic visualisation was performed with DAB Peroxidase substrate kit (Vector laboratories) according to the manufacturer's instructions. Slides were counterstained with Mayer's HTX Haematoxylin (Histolab, Gothenburg, Sweden) for 30 seconds, dehydrated and permanently mounted (Vecta Mount, Vector laboratories). A similar procedure was performed for CD45 (2B11+PD7126, mouse monoclonal DAKO Agilent; dilution 1:100, detection with secondary ImmPRESS HRP Anti-mouse IgG) and HER2 (D8F12, rabbit monoclonal; CellSignalling dilution 1:50, detection with secondary ImmPRESS HRP Anti-rabbit IgG (Vector Laboratories)).

Standard immunohistochemistry staining on non-ISS sequenced sections was performed for the following proteins: SM-MHC (BioCare Medical LLC, cat# CM 420B, mouse monoclonal, clone: SMMS-1, dilution 1:100), P63 (BioCare Medical LLC, cat# CM 163C, mouse monoclonal, clone: BC4A4, dilution 1:150), PR (DakoCytomation, cat# M3569, mouse monoclonal, clone: PgR636, dilution 1:75), Ki67 (DakoCytomation, cat# M7240, mouse monoclonal, clone: MIB-1, dilution 1:400) and PTEN (Abcam Anti-PTEN antibody [EPR22636-122] (ab267787) 1:500 pc). Secondary antibody detection was performed using polymer-M (Labelled polymer-HRP anti-mouse: DakoCytomation, Code: K4007) for PR, Ki67, P63 and PTEN (EPR22636-122) and poly-AP-M (Poly-AP anti-mouse IgG : Leica, cat# PV6110) for SMMHC and polymer-M (Labelled polymer-HRP anti-mouse: DakoCytomation, Code: K4007) for PR, Ki67, P63 and PTEN (EPR22636-122) and poly-AP-M (Poly-AP anti-mouse IgG : Leica, cat# PV6110) for SMMHC.

In selected regions (with LCM-WGS validated subclone compositions) the total number and number of positive stained nuclei (Ki-67, PTEN and PR antibodies) were counted

using Qupath digital software (Version 0.3.0) [Bankhead et al. 2017].

A.3. Bulk tissue sequencing

To generate DNA and RNA sequence data 10 x 10-20um serial slices of tissue blocks were pooled and homogenised followed by nucleic acid extraction. Short insert whole genome and targeted capture paired-end libraries and 350bp poly-A selected RNA libraries were created, flow cells prepared and sequencing clusters generated according to Illumina protocols [Kozarewa et al. 2009]. 100 base whole genome sequence data, 150 base targeted capture genomic data and 75 base RNA sequence data was generated using Illumina HiSeq 2000® genome analysers. Genomic and RNA sequence data were mapped to the reference human genome (GRCh37 Ensembl 58) using Burrows Wheeler Aligner [Heng Li et al. 2009] and TopHat (v1.3.3), respectively. Original data associated with the prior publication datasets from which these samples derive are deposited in the European Genome Phenome Archive (EGA) with the following accessions: EGAD00001002696 and EGAD00001000898.

Somatic Mutation Identification

Genome-wide somatic substitutions were called using CaVEMan (Cancer Variants Through Expectation Maximisation). Rearrangements were identified from discordantly mapping reads using BRASSII (BReakpoint AnalySiS).

A.4. Inferring subclone composition and evolutionary histories from bulk genomic data

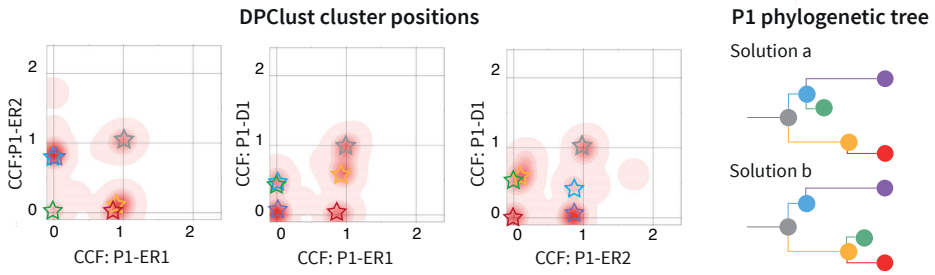
Subclone composition was determined from multi-region WGS somatic substitution and copy number data as reported in the original publications Yates et al. [2015] and Yates2017-xc. The approach detects a finite number of subclones each with defined parameters including the number of mutations within that subclone and the fraction of cells from each sample that contain those subclone specific mutations (Box 3).

A.4.1. cancer cell fractions

For bulk genomic sequence data the variant allele frequency (VAF) of each mutation, i, is calculated from the number of reads reporting the mutant r_{mut} and the number of reference, or wild-type reads r_{wt} is calculated as

$$VAF_i = \frac{r_{mut,i}}{r_{mut,i} + r_{wt,i}} \tag{A.1}$$

However, the VAF cannot be used directly to infer subclonal architecture because in addition to reflecting the proportion of cells that carry that mutation it is also influenced



Appendix Figure A.1: Phylogeny reconstruction from multi-region WGS data of P1. Density plots of WGS derived point mutation, cancer cell fraction (CCF) estimates from pairs of samples (see Supplementary Methods for details). Mutation clusters are denoted by coloured stars. The two phylogenetic tree solutions most compatible with the mutation cluster CCFs are presented alongside their respective inferred P1-green genotypes

by the tumour purity and the local copy number – with the latter encompassing both the total number of alleles and the number of alleles that carry the mutation [Dentro et al. 2017]. Consequently, mutations from the same subclone may have different VAFs reflecting their different copy number states, also termed their "multiplicity". To allow a comparison of mutations across related samples with different copy number profiles and tumour purity levels we derive the mutation copy number. For mutation i, the mutation copy number n_i is obtained as

$$n_i = VAF_i \frac{1}{\rho} [\rho n_{locus,t} + n_{locus,n}(1-\rho)] \tag{A.2}$$

Where ρ is tumour purity, and $n_{locus,t}$ and $n_{locus,n}$ are the locus specific total copy numbers in the tumour and normal sample respectively. Subclonal copy-number changes and tumour purity are defined using the Battenberg algorithm as previously described. Confidence intervals for mutation copy numbers are generated using bootstrap resampling of mutant and wildtype reads from each mutant locus (10,000 times) and applying the above formula to resampled reads.

For mutations that are present on multiple alleles arising through duplication (multiplicity $\neq 1$) or in regions of subclonal copy number change, the mutation copy number may not accurately reflect the fraction of cells in which the mutation resides. To group mutations by the number of cells containing those mutations, the number of chromosomes bearing a mutation, nchr, must be determined. For each mutation within a region of amplified major copy number C, the observed number of mutant and wild-type reads are compared to the expected VAF_i that would result from the mutation being present in $1, 2 \ldots C$ chromosome copies also allowing for non-integer values in regions with subclonal copy number values, assuming a binomial distribution. The value of nchr is determined to be that with the maximum likelihood. The fraction of cells bearing mutation i, termed the CCF can then be calculated as

$$CCF_i = \frac{n_i}{n_{chr}} \tag{A.3}$$

Applying this approach to cancers that are believed to derive from a common ancestor, one would therefore expect to identify some clonal mutations that are present in all cancer cells and at a certain multiplicity result in a CCF around 1.

A.4.2. Multidimensional mutation clustering

A Bayesian Dirichlet Process was used to model clusters of clonal and subclonal point mutations to permit the inference of the number of subclones and the fraction of cells (CCF) within each subclone (Appendix Figure A.1 and Box 3). The original model was developed for single samples [Nik-Zainal, Van Loo et al. 2012]. Within this model the number of reads y, bearing the ith mutation is drawn from a binomial distribution as follows

$$y_i \sim \text{Binom}(\pi_i, N_i) \tag{A.4}$$
$$\pi_i \sim \mathcal{DP}(\alpha P_0) \tag{A.5}$$

Where N_i is the number of mutant reads and ζ_i is the expected fraction of reads that would report a mutation if it was present in 100% of cells (at that specific locus given the copy number state and tumour purity). The fraction of tumour cells carrying mutation i, π_i, is modelled as arising from a Dirichlet process (DP) with a given probability distribution P_0 and a dispersion parameter α. The approach allows co-estimation of the number of cellular populations and their properties using the stick-breaking representation of the DP as described in Dentro et al. [2017].

To accommodate multi-region samples, the model is extended into multiple dimensions as previously described [Bolli et al. 2014; Yates et al. 2015]. Essentially, the number of mutant reads obtained from different tumour region samples are modelled as independent binomial distributions with independent π drawn with a Dirichlet process. Gibbs sampling is used to estimate posterior distributions of the features of interest. The Markov chain is run for 10,000 iterations (excluding the first 2,000 iterations). The median densities for CCFs is estimated using a Gaussian kernel implemented in R (stats and kernSmooth libraries). The clustering procedure is implemented in the DPClust software package. Version 2.2.8 was used to obtain the analysis presented.

In P2 (samples P2-TN1, P2-TN2, P2-LN1) the final DPClust solution identified a single subclone within the lymph node sample (P2-LN1) however there was evidence supporting the possible existence of a subclone albeit with lower confidence, understood in the context of low tumour purity (~11%). This is most convincing from copy number analysis that identified multiple sub-chromosomal copy number changes (chromosome 1,9, 11,12,20) in addition to a 50% subclone at the telomeric portion of 17q in association with the *HER2* amplifying breakage fusion bridge event (Figure 4.10). This was further suppor-

ted by a build-up of posterior density in P2-LN1, indicating that across some iterations DPClust has explored the presence of a subclone in this sample was explored, although ultimately it decided the subclone wasn't required to explain the observed point mutation data.

A.4.3. Principles of phylogenetic tree construction

To infer the evolutionary relationship between subclones, phylogenetic trees were constructed using the "pigeonhole principle" which states that if there are n objects that are placed in m containers if $n > m$ then at least one container must contain more than one object (Box 3). Extending this to subclonal reconstruction we can appreciate that the sum of subclone CCFs cannot exceed the CCF of their ancestor. For example, if there are 3 subclones with CCFs of 1, 0.9 and 0.5 then there must be a linear evolution – as 1 + 0.9 > 1, CCF subclone 0.9 must be a descendant of the CCF subclone 1; furthermore, as 0.9 + 0.5 > 1, CCF subclone 0.5 must be a descendant of the CCF subclone 0.9. In contrast one may see the possibility of branching evolution where the sum of subclone populations is less than 1. For example, 3 subclones with CCFs of 1, 0.5 and 0.4 could arise either through linear evolution: 1 > 0.5 > 0.4 or along separate ancestral lines because 0.5 + 0.4 < 1. When applying these principles across multi-region samples the compatible underlying tree structures is usually greatly restricted because of the requirement that the same tree structure must be compatible across all samples. Applying these principles one can identify the phylogenetic tree structure(s) most compatible with the underlying data (Appendix Figure A.1).

A.5. BaSISS and ISS protocols

A.5.1. Tissue specimens

Serial 10µm tissue sections were cut from fresh frozen tumour tissue blocks mounted on superfrost slides and sent from DFCI to ScilifeLab Stockholm, Sweden where ISS and IHC was performed under Karolinska Institutes rules for the handling of blood and other human sample material, reference number 1-31/2019 (with a local HUMRA risk assessment form). Tissue blocks included the same ones used for the WGS and RNAseq experiments and additional tumour blocks were identified where applicable.

A.5.2. Padlock probe design

The padlock probe consists of a single stranded DNA oligo with two target recognition arms in the 3' and 5' end of the oligo, enabling circularization of the padlock probe upon target hybridization. Each target recognition arm typically has a length of 15-25 nucleotides. The two arms are interspaced with a linker sequence comprising a 20-nucleotide anchor primer sequence and a target-specific 4 nucleotide barcode followed by a 5 nuc-

leotide stabilising sequence for sequencing-by-ligation. The used barcodes were selected in such a way that they differ in at least two positions.

In total, three padlock probe gene panels were used, marker/oncology gene-, mutation- (one specifically designed for each case) and immune- expression panels. Padlock probes were ordered as ultramer DNA oligos from Integrated DNA Technologies (IDT, Leuven, Belgium) with 5'-phosphorylation modification and were reconstituted in Tris/EDTA buffer.

A.5.3. Mutation panel design

Two separate case-specific mutation panels were designed as detailed in Table A.1. For wildtype- and mutation-specific padlock probes, the sequence upstream of the mutated site resembles the 3' arm of the padlock probe, with the wildtype or mutant nucleotide at the 3' end, whereas the sequence downstream of the mutated site resembles the 5' arm. The target recognition arm lengths were adjusted to have a similar melting temperature of ~55℃. For each mutation site, one wild-type and one mutation-specific padlock probe was designed. For P1, in addition to the anchor and barcode sequences a 20-nucleotide sequence used for a hybridization cycle was added in the padlock probe linker sequence. Specific primers were used for the in situ reverse transcription and were designed as the reverse complement sequence of the 5' target recognition arm. Primers were ordered as DNA oligos from Integrated DNA Technologies (IDT) and were reconstituted in Tris/EDTA buffer. In P1, we included 3 mutations that were not assigned to a cluster/ branch by DPclust of WGS data. In accordance with this, BaSISS signals reporting these mutations were found to be non-specific and uninformative, probably relating to them existing at loci with variable copy number states, and they were dropped from further downstream analyses.

Table A.1: BaSISS padlock probes for the P1 case. Padlock probes targeted specific sequence variants based on WGS and DPclust clone discovery results. The last column indicates if the probe was included in the clone mapping analysis (Chapter 2). CDS = coding sequence; HK = housekeeping

Branch	Gene	Target Name	CDS	Type	Barcode	Included?
P1-blue	*CKAP5*	CKAP5mut	c.1895G>A	mut.	ACCGG	Y
P1-blue	*CKAP5*	CKAP5wt	c.1895G>A	wt..	CCACC	Y
P1-green	*KIAA0652*	KIAA0652mut	c.176C>G	mut.	GACAA	Y
P1-green	*KIAA0652*	KIAA0652wt	c.176C>G	wt..	ACCAC	Y
P1-grey	*AP3B2*	AP3B22mut	c.1852+5G>C	mut.	CCCAC	Y
P1-grey	*AP3B2*	AP3B22wt	c.1852+5G>C	wt..	AAGAG	Y
P1-grey	*ARHGEF28*	ARHGEF28mut	c.1084C>G	mut.	AGAAG	Y
P1-grey	*ARHGEF28*	ARHGEF28wt	c.1084C>G	wt..	GCGGC	Y
P1-grey	*CCDC105*	CCDC105mut	c.1227C>A	mut.	GCAAA	Y
P1-grey	*CCDC105*	CCDC105wt	c.1227C>A	wt.	GAACA	Y
P1-grey	*CREBBP*	CREBBPmut	c.2827C>T	mut.	AGGCC	Y

P1-grey	*CREBBP*	CREBBPwt	c.2827C>T	wt.	GGGAA	Y
P1-grey	*DENND1A*	DENND1Amut	c.2960C>T	mut.	AACAA	Y
P1-grey	*DENND1A*	DENND1Awt	c.2960C>T	wt.	CAAGG	Y
P1-grey	*KCNT1*	KCNT1mut	c.3409C>A	mut.	CACAA	Y
P1-grey	*KCNT1*	KCNT1wt	c.3409C>A	wt.	GAAGG	Y
P1-grey	*NOB1*	NOB1mut	c.503A>G	mut.	GCACC	Y
P1-grey	*NOB1*	NOB1wt	c.503A>G	wt.	AACCC	Y
P1-grey	*OXSM*	OXSMmut	c.1268A>G	mut.	AGAGA	Y
P1-grey	*OXSM*	OXSMwt	c.1268A>G	wt.	ACGGC	Y
P1-grey	*PQLC2*	PQLC2mut	c.627G>A	mut.	CGGAA	Y
P1-grey	*PQLC2*	PQLC2wt	c.627G>A	wt.	CGAGA	Y
P1-grey	*RELA*	RELAmut	c.341A>T	mut.	CCCGG	Y
P1-grey	*RELA*	RELAwt	c.341A>T	wt.	AAGGA	Y
P1-grey	*SF3B1*	SF3B1mut	c.2098A>G	mut.	AGGAA	Y
P1-grey	*SF3B1*	SF3B1wt	c.2098A>G	wt.	CGCCG	Y
P1-grey	*STUB1*	STUB1mut	c.478G>A	mut.	ACACC	Y
P1-grey	*STUB1*	STUB1wt	c.478G>A	wt.	GAGAG	Y
P1-orange	*LRP1B*	LRP1Bmut	c.10533G>A	mut.	GGCCG	Y
P1-orange	*LRP1B*	LRP1Bwt	c.10533G>A	wt.	GGAAG	Y
P1-purple	*AMZ1*	AMZ1mut	c.208C>T	mut.	CCGGC	Y
P1-purple	*AMZ1*	AMZ1wt	c.208C>T	wt.	AAAGG	Y
P1-purple	*EMILIN2*	EMILIN2mut	c.1896T>C	mut.	ACAAA	Y
P1-purple	*EMILIN2*	EMILIN2wt	c.1896T>C	wt.	AAACA	Y
P1-purple	*FZD4*	FZD4mut	c.356G>C	mut.	GCCGG	Y
P1-purple	*FZD4*	FZD4wt	c.356G>C	wt.	CAGGA	Y
P1-purple	*PTEN*	PTEN1mut	c.388C>T	mut.	CGGCC	Y
P1-purple	*PTEN*	PTEN1wt	c.388C>T	wt.	CGCGC	Y
P1-purple	*TMEM8A*	TMEM8Amut	c.735G>A	mut.	GCCAC	Y
P1-purple	*TMEM8A*	TMEM8Awt	c.735G>A	wt.	CAGAG	Y
not assigned	*AP3*	ACTB	–	HK	GCGCG	N
not assigned	*FGFR1*	FGFR1expNew	–	exp.	CAAAC	N
not assigned	*DSEL*	DSELmut	c.1701G>A	mut.	CCAAA	Y
not assigned	*DSEL*	DSELwt	c.1701G>A	wt.	CAACA	Y
not assigned	*KIF14*	KIF14mut	c.573G>A	mut.	AGCGC	Y
not assigned	*KIF14*	KIF14wt	c.573G>A	wt.	GGAGA	Y
not assigned	*PLXNA2*	PLXNA2mut	c.100G>A	mut.	GGCGC	Y
not assigned	*PLXNA2*	PLXNA2wt	c.100G>A	wt.	CGAAG	Y

Table A.2: BaSISS padlock probes for the P2 case. Padlock probes targeted specific sequence variants, fusions, and gene expressions based on WGS.

Sample	Gene	Target Name	RNA	Type	Barcode	Included?
		Single Nucleotide Variants				
All	*COX19*	COX19wt	r.394c>g	wt.	GGCC	Y

All	*COX19*	COX19mut	-	mut.	GGAA	Y
P2-LN1	*ERBB2*	ERBB2wt	r.4370g>u	wt.	GCGC	Y
P2-LN1	*ERBB2*	ERBB2mut	-	mut.	GCCG	Y
-	*RPL37A*	RPL37Awt	r.985g>c	wt.	GAGA	Y
P2-TN1	*RPL37A*	RPL37Amut	-	mut.	GAAG	Y
-	*TP53*	TP53wt	r.728g>u	wt.	CGGC	Y
All	*TP53*	TP53mut	-	mut.	AAGG	Y
-	*PLXNA1*	PLXNA1wt	r.7031c>u	wt.	CCGG	Y
P2-TN1/TN2	*PLXNA1*	PLXNA1mut	-	mut.	AACA	Y
		Fusions				
P2-LN1	*CACNB1*	CACNB1ex9-intr10	-	Splice	AGAG	N
P2-LN1	*CACNB1*	CACNB1intr10-ex10	-	Splice	ACCC	Y
-	*CACNB1*	CACNB1ex9-ex10wt	-	wt.	ACAA	Y
-	*CARTPT-PID1*	CARTPT-PID1	-	Fusion	CCAC	N
-	*ZNF652:SNHG5*	ZNF652:SNHG5	-	Fusion	CACC	N
		Gene Expression				
P2-TN2/TN1	*CDK6*	CDK6exp	-	Amplified	CAAA	N
P2-LN1	*ERBB2*	ERBB2exp	-	Amplified	AGGA	Y
-	*ACTB*	ACTB	-	HK	CGCG	N

A.5.4. Immune panel design

Large scale probe design was facilitated using an in-house Python software package as described previously [Qian et al. 2020] which utilises ClustalW and BLAST+ to ensure probe specificity. Each padlock probe of the immune panel was designed to contain two 20 nucleotide long target recognition arms. Only target fragments with melting temperature between 65℃ and 75℃ were considered. Probes were selected aiming to obtain a distribution along the whole length of the transcript.

Overall, five padlock probes were selected per target gene with the exception of *HLA-DRB1*, where only two specific probes could be designed. For the T cell specific genes *CD8A*, *FOXP3*, *EOMES*, *CD4* and *IFNG* 20 probes were selected in order to increase the detection efficiency for these target genes.

A combination of random decamer primers (IDT, Leuven Belgium) and specific primers were used for in situ reverse transcription. Specific primers were designed to hybridise to the mRNA 15-20 nucleotides downstream of the target sequence. Primers were ordered as DNA oligos from Integrated DNA Technologies (IDT) and were reconstituted in Tris/EDTA buffer.

Target genes for the immune panel were selected to cover a broad range of immune cell subtype markers with special emphasis on T cell subsets and their regulation. For the complete list of the immune panel genes, refer to the original publication [Lomakin et al. 2022].

A.5.5. Oncology gene panel design

The oncology gene panel has previously been published and includes genes involved in proliferation, EMT, invasiveness, stemness, angiogenesis as well as genes for breast cancer subtyping and oncotypeDX recurrence scoring [Svedlund et al. 2019]. The target recognition arms were designed to capture most splice variants of the gene transcripts and blasted to confirm their specificity. Each target recognition sequence had a GC content of 50-55% and a melting temperature of ~55℃. In this older design, the panel included one padlock probe per gene target and the barcodes used were only differing in one position. For the marker gene panel, random decamer primers were used for in-situ reverse transcription (IDT, Leuven Belgium).

A.5.6. *in situ* sequencing

ISS was performed as described by Ke et al. [2013] and modified according to Svedlund et al. [2019], and was used to spatially resolve oncology panel-, mutation panel- and immune panel- gene expression profiles on consecutive sections from the breast tumour tissue blocks.

The ISS library preparation and sequencing is described in detail at protocols.io.bb2giqbw, the steps in the protocol that were modified for breast cancer tissues are indicated below. In brief, library preparation included fixation of tissue sections with 4% PFA for 30 min (step 2) followed by permeabilized with 0.1 mg/ml pepsin (Sigma) in 0.1 M HCl 37℃ for 90 s (step 4). SecureSeal™ reaction chambers were mounted on top of the tissues (Grace Biolabs, Bend, United States) and cDNA was synthesised in situ using specific DNA primers (125nM for mutation panel, 5nM for immune panel) and/or random decamer primers (5µM for immune and marker gene panels) (step 10) (IDT, Leuven Belgium, sequences are listed in the Supplementary Table 1 in Lomakin et al. [2022]). Rnase H was used to generate single-stranded cDNA that the padlock probes could hybridise to. Hybridized padlock probes (10nM of each in immune panel, onM of each in mutation and marker gene panel, step 15) (IDT, Leuven Belgium, sequences are listed in the Supplementary Table 1 in Lomakin et al. [2022]) were ligated using T_{th} ligase, a highly specific DNA ligase that can discriminate correct base-pairing at the single nucleotide level. Only completely target-complementary padlock probes become ligated, forming closed circles that could then be amplified through rolling circle amplification (RCA). Of note, for P1, the experimental conditions for the first replica of ISS differed slightly with a diverse Phi29 buffer (Thermo Fisher 10X reaction buffer: 330 mM Tris-acetate (pH 7.9 at 37℃), 100 mM Mg-acetate, 660 mM K-acetate, 1% Tween 20 and 10 mM DTT) and no Exonuclease 1 in the rolling circle amplification step.

The four-five nucleotide target-specific barcodes included in the padlock probe linker sequence were clonally amplified in the RCA products allowing identification through sequencing by ligation of anchor primer and fluorophore-labelled interrogation probes [Ke et al. 2013]. Nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI). The target-specific barcodes were sequenced with four sequencing and imaging rounds. For

case P1, the mutation panel was sequenced with one hybridization cycle in addition to the four sequencing by ligation cycles. After the ISS analysis, the tissue sections were stained with PanCK, CD45 or HER2 antibody (Appendix A.2).

A.5.7. Imaging

Images were acquired with an automated Zeiss Axioplan II epifluorescence microscope (Zeiss, Oberkochen, Germany) using a z-stack of 0.49µm× 11 and a tile overlap of 1%. Images were scanned with a 20X objective. For the first base sequenced, the exposure times were calibrated so that the signal intensity values were similar for all sequencing channels (A–Cy5, G-Cy3, C-Texas Red and T-AF488), the calibrated exposure times were then kept constant for all remaining sequencing cycles. Orthogonal projections and stitching of tiles were done with the ZEN software (Zeiss).

Image stitching

From a total of 51 image sets, 43 were stitched with Carl-Zeiss ZEN software (version 3.1), and the other 8 failed image sets were stitched using BigStitcher (version 0.9) [Hörl et al. 2019].

A.6. Image data processing

A.6.1. Image registration

The registration across imaging cycles was performed in two steps: affine registration on DAPI channel and subsequently local warping on anchor channel. For both steps we used algorithms provided in libraries OpenCV-contrib (version 4.3.0) [BRADSKI et al. 2000] and scikit-image (version 0.17) [Walt et al. 2014]. In all imaging cycles, before the registration, both DAPI and anchor channels were maximum intensity projected across the image z-stack.

During the affine registration step, we coarsely align images of all cycles to the first one based on the DAPI channel. Firstly, we detect key points in the images of each cycle using the FAST feature detector. Secondly, for each key point, its surrounding area is described with histograms of oriented gradients using the DAISY feature descriptor. After that, using the key points and their descriptors, the FLANN-based matcher finds correspondences between pairs of key points from reference and moving images and filters out unreliable points. Lastly, the remaining key points are processed using the RANSAC-based algorithm that aligns them and estimates affine transformation parameters with 4 degrees of freedom.

The second registration step aligns imaging cycles sequentially using the anchor channel (fluorophore Cy7). We applied the Farneback optical flow algorithm to achieve more accurate registration by warping the images locally, so that RNA spots of different channels can be better aligned despite the presence of nuclei swelling, imperfect stitching and

sample distortion. In both steps, we optimised the algorithm by performing computation on the tiled images to reduce memory consumption and accelerate the transformation parameters estimation.

A.6.2. Serial tissue image signal alignment

Analysis of sample P2-LN required IHC signal projection performed on a consecutive slide back to BaSISS slide. To achieve this, we performed a spline-based elastic registration implemented in ImageJ package UnwarpJ [Arganda-Carreras et al. 2008].

A.6.3. ISS signal deconvolution

After registration of images from different sequencing rounds, we locate RNA spots by applying the circular Hough transform to the reference anchor channel of the first round, which is implemented in MATLAB's function 'imfindcircles'. At the detected coordinates, image values are extracted from top-hat filtered coding channels across all sequencing rounds. We then perform decoding of the extracted image values via a Gaussian Mixture Model, where each mixture corresponds to one of the possible barcodes encoded via an experimental design. Finally, once the mixture model is fitted to the extracted image values, each detected spot is assigned to the most likely barcode [Gataric et al. 2021]. In addition to on-target barcodes, there is an infeasible class which represents RNA spots to which barcodes could not be assigned. In Supplementary Table 4 we present QC metric of our datasets as a proportions of RNA spots with assigned barcodes ($p > 0.6$) to the total number of spots.

A.7. LCM-WGS Validation

Residual frozen tissue blocks from P1 samples P1-D1, P1-D2, P1-ER1 and P1-ER2 were accessed to perform laser capture microdissection (LCM) and low input library whole genome sequencing as previously reported [Ellis et al. 2021]. Ten micrometre thick sections were generated at -20 ℃ and mounted on poly-ethylene naphtholate (PEN)-membrane slides (Leica), fixed with 70% ethanol for 2 minutes followed by serial immersion in the following: deionised water (1 minute), Gill's haematoxylin (10 seconds), tap water (20 seconds, twice), eosin (5 seconds), tap water (20 seconds), ethanol 70% (20 seconds, twice), ethanol 100% (20 seconds, twice), xylene (20 seconds, twice).

Using a laser-capture microscope (Leica LMD7), breast tissue architecture in relation to existing mapped clones (where feasible) was first visualised, then regions of approximately 13,000-180,000μm^2 were dissected (power 55, aperture 2) and collected into the individual wells of a 96-well plate. Samples were then processed as described by Ellis et al. [2021]. Briefly, individual steps include cell lysis and digestion using Arcturus PicoPure Protease buffer and thermal cycling in a sealed plate (60 ℃ for 3 h, 75 ℃ for 30 min, hold at 4 ℃), gDNA purification using Agencourt AMPure beads, enzymatic DNA library construction (NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs,

cat. no. E7805L)), adapter ligation and amplification. Genomic libraries with concentrations of >5ng/µl DNA (total volume = 20µl) were selected for NovaSeq 6000 paired end sequencing. A bulk whole genome library was also generated from whole blood derived DNA and sequenced in the same way. As described above for bulk WGS analysis, point mutations were called using CaVEMan and copy number using ASCAT and Battenberg algorithms and mutation clusters identified using DPclust v2.2.8 (Supplementary Data Table 5 in Lomakin et al. [2022]).

A.8. Mutation timing estimates

We report that there is evidence of early divergence of the subclones detected in P1 (prior to 50% of in m time). Evolutionary timing estimates can be sought for P1 but are not attempted for P2 due to relatively low purity. Based on the assumption of a relatively constant mutation rate within a given genetic lineage, subclone divergence as a percentage of evolutionary time can be calculated from the number of mutations accumulated up to the branching point (branch lengths) divided by the longest series of branches within that lineage. Mutation rates particularly in cancers might be greatly altered by hypermutator states, however, we confirmed that in P1 virtually all mutations in this case are derived from clock-like mutational processes using our previously published approaches (Signature 1, 5 and 40) (Supplementary Table 3, in Lomakin et al. [2022]) [Nik-Zainal et al. 2016].

Supplementary Information to Chapter 2

B

Declaration

This supplementary chapter presents tables that complement Chapter 2. All content derives from the supplementary materials of Lomakin et al. [2022], although with minor stylistic modifications.

Table B.1: Genotype matrix for P1. The genotype matrix for allele copy numbers serves a critical role in inferring the clone map. This matrix contains numerical values that signify the presumed copy number state of allelic variants in each clone. It is derived as outlined in Appendix A.4

Locus	P1-grey	P1-green	P1-purple	P1-blue	P1-red	P1-orange	wt
SF3B1, mut/wt	1 / 2	1 / 2	1 / 2	1 / 2	1 / 2	1 / 2	0 / 2
STUB1, mut/wt	1 / 2	1 / 2	1 / 2	1 / 2	1 / 2	1 / 2	0 / 2
CREBBP, mut/wt	1 / 2	1 / 2	1 / 2	1 / 2	1 / 2	1 / 2	0 / 2
ARHGEF28, mut/wt	1 / 2	1 / 2	1 / 2	1 / 2	1 / 2	1 / 2	0 / 2
KIAA0652, mut/wt	0 / 2	1 / 1	0 / 2	0 / 2	0 / 2	0 / 2	0 / 2
OXSM, mut/wt	1 / 2	1 / 1	1 / 1	1 / 1	1 / 2	1 / 2	0 / 2
CKAP5, mut/wt	0 / 2	1 / 1	1 / 1	1 / 1	0 / 2	0 / 2	0 / 2
DENND1A, mut/wt	0 / 2	0 / 2	0 / 2	0 / 2	1 / 0	1 / 0	0 / 2
NOB1, mut/wt	1 / 2	1 / 2	1 / 1	1 / 2	1 / 0	1 / 0	0 / 2
RELA, mut/wt	0 / 0	0 / 0	0 / 0	0 / 0	1 / 0	1 / 0	0 / 0
PLXNA2, mut/wt	1 / 2	1 / 1	1 / 1	1 / 2	1 / 2	1 / 2	0 / 2
PQLC2, mut/wt	0 / 2	1 / 1	1 / 1	1 / 2	0 / 2	0 / 2	0 / 2
LRP1B, mut/wt	0 / 2	0 / 2	1 / 1	0 / 2	0 / 2	0 / 2	0 / 2
PTEN1, mut/wt	0 / 2	0 / 2	0 / 2	0 / 2	1 / 0	0 / 2	0 / 2
PTEN2, mut/wt	0 / 2	0 / 2	1 / 0	0 / 2	0 / 2	0 / 2	0 / 2
TMEM8A, mut/wt	0 / 2	0 / 2	1 / 1	0 / 2	0 / 2	0 / 2	0 / 2
DSEL, mut/wt	0 / 3	0 / 3	1 / 3	0 / 3	0 / 4	0 / 4	0 / 2
FGFR1exp	3	3	4	3	16	12	2
KIF14, mut/wt	0 / 1	0 / 1	1 / 1	0 / 1	0 / 2	0 / 2	0 / 2
AMZ1, mut/wt	0 / 2	0 / 2	1 / 1	0 / 2	0 / 2	0 / 2	0 / 2
KCNT1, mut/wt	1 / 2	1 / 2	1 / 1	1 / 2	1 / 2	1 / 2	0 / 2
FZD4, mut/wt	0 / 1	0 / 1	1 / 1	0 / 1	0 / 2	0 / 2	0 / 2

AP3B22, mut/wt	1 / 2	1 / 2	1 / 1	1 / 2	1 / 2	1 / 2	0 / 2
EMILIN2, mut/wt	0 / 2	0 / 2	1 / 1	0 / 2	0 / 2	0 / 2	0 / 2
CCDC105, mut/wt	1 / 2	1 / 2	1 / 1	1 / 2	1 / 2	1 / 2	0 / 2
ZNF468, mut/wt	1 / 2	1 / 2	1 / 1	1 / 2	1 / 2	1 / 2	0 / 2

Table B.2: Genotype matrix for P2. The genotype matrix for allele copy numbers serves a critical role in inferring the clone map. This matrix contains numerical values that signify the presumed copy number state of allelic variants in each clone. It is derived as outlined in Appendix A.4

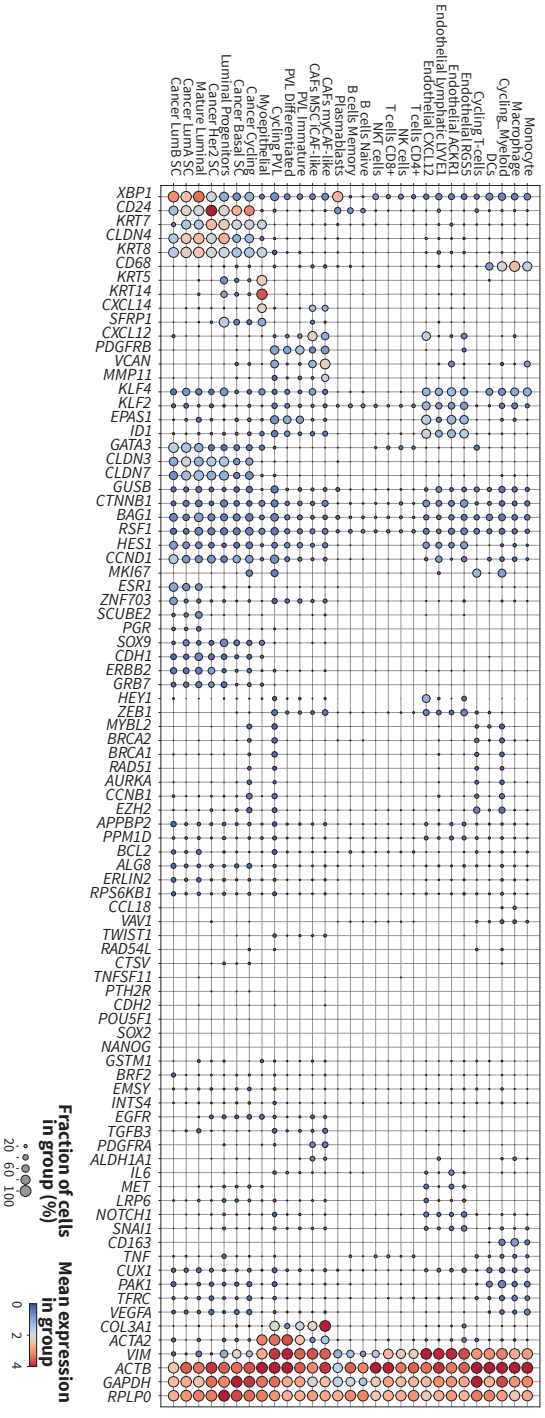
Locus	P2-blue	P2-green	P2-orange	P2-purple	wt
TP53, mut/wt	2 / 0	2 / 0	2 / 0	2 / 0	0 / 2
COX19, mut/wt	1 / 2	2 / 2	2 / 2	2 / 3	0 / 2
PLXNA1, mut/wt	1 / 4	1 / 4	1 / 4	1 / 3	0 / 2
RPL37A, mut/wt	0 / 4	0 / 4	0 / 4	1 / 4	0 / 2
ERBB2, mut/wt	0 / 2	1 / 5	30 / 150	0 / 2	0 / 2
ERBB2exp	2	52	52	2	2
CACNB1ex9-ex10wt	3	4	80	3	2
CACNB1intr10-ex10	0	0	20	0	0

Supplementary Information to Chapter 3

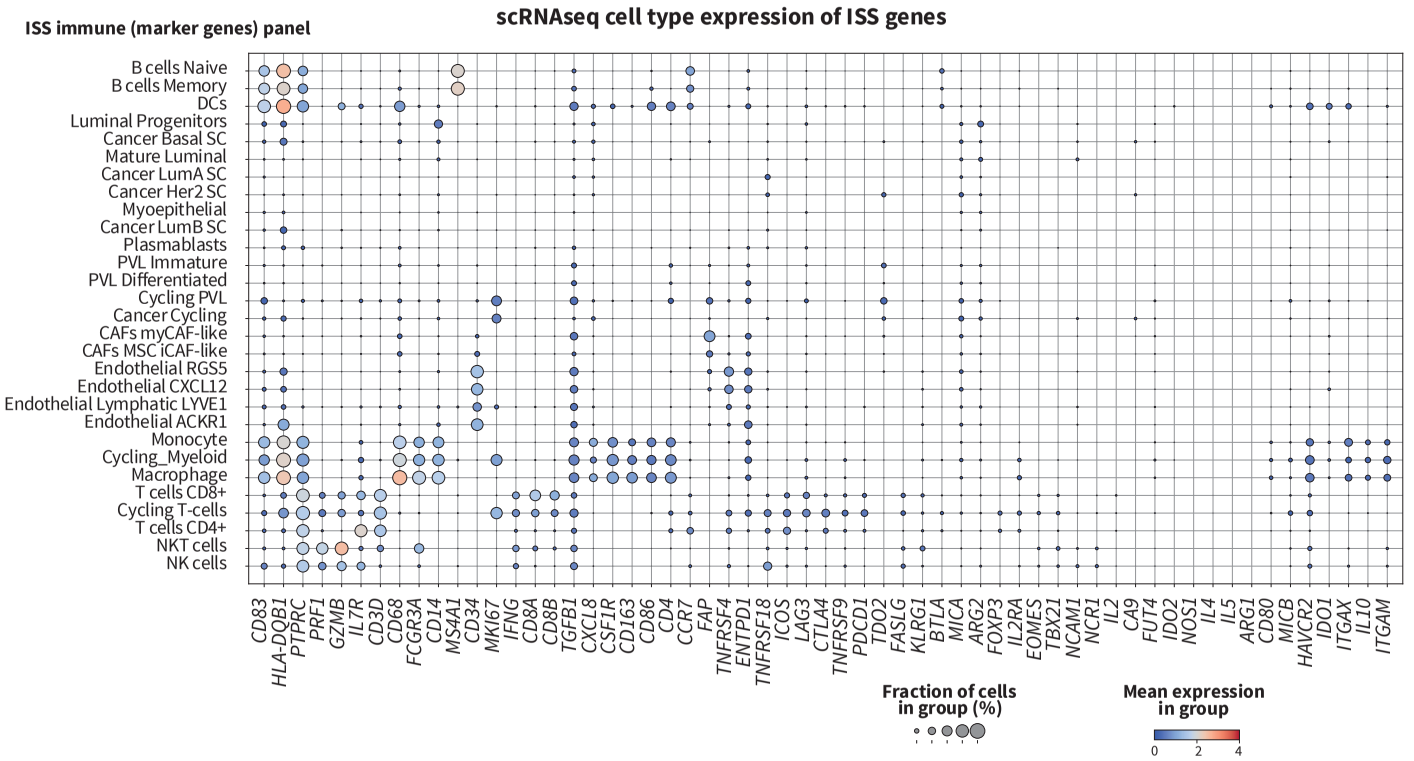
C

Declaration

This supplementary chapter presents tables and figures that complement Chapter 3. All content derives from the supplementary materials of Lomakin et al. [2022], although with minor stylistic modifications.



Appendix Figure C.1: Mean expression of the genes used in ISS oncology panel. This table was calculated from the breast cancer scRNA-seq reference (derived from S. Z. Wu et al. [2021]) to aid interpretation of the observed ISS signal distribution.



Appendix Figure C.2: Mean expression of the genes used in ISS immune panel. This table was calculated from the breast cancer scRNA-seq reference (derived from S. Z. Wu et al. [2021]) to aid interpretation of the observed ISS signal distribution.

Supplementary Information to Chapter 4

D

Declaration

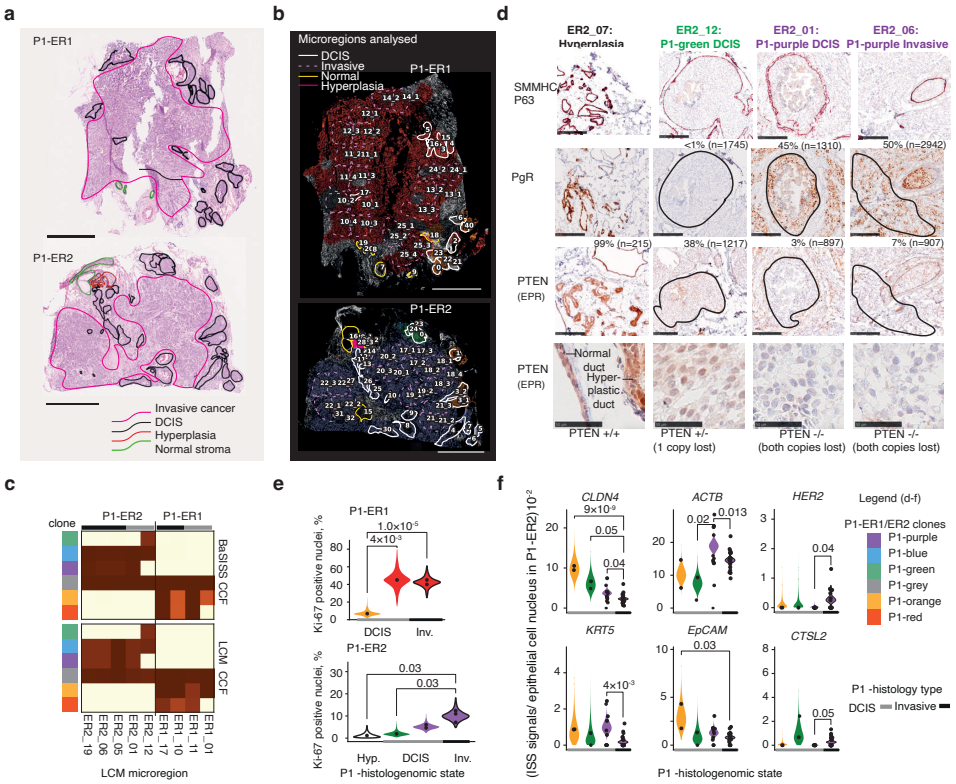
This supplementary chapter presents tables and figures that complement Chapter 4. All content derives from the supplementary materials of Lomakin et al. [2022], although with minor stylistic modifications.

Table D.1: Patient clinical description. Clinical information for the studied breast cancer cases

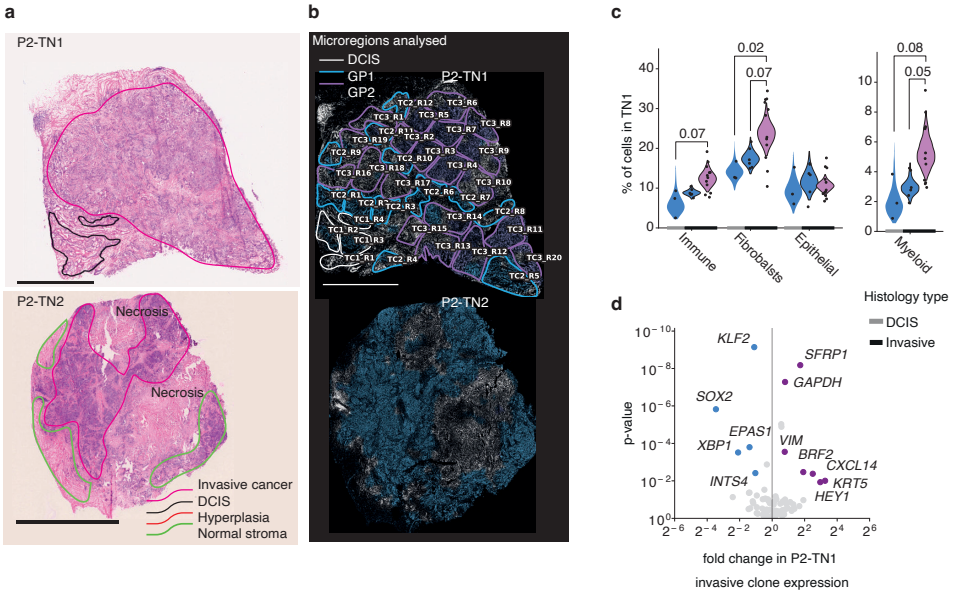
Parameter	P1 case	P2 case
Manuscript ID	P1	P2
Case	PD9694	PD14780
Sex	Female	Female
Age at Diagnosis (Years)	37	66
Children	Unknown	Yes
Maximum Tumour Size (mm)	26	Not Recorded
Axillary Nodes	Micromets (1LN)	3+
Metastasis Status at Diagnosis	M0	M0
Histology	Invasive Carcinoma: No Special Type	Invasive Carcinoma: No Special Type
In Situ Component	Extensive Intermixed DCIS	No
Multifocal (Number of Invasive Foci)	Yes (2)	Yes (3)
Grade Primary	2	Unknown
ER Status Primary	Positive	Negative
PR Status Primary	Positive	Negative
HER2 Primary (IHC)	Negative	Negative

Table D.2: Sample clinical description. Clinical data related to the studied breast cancer samples

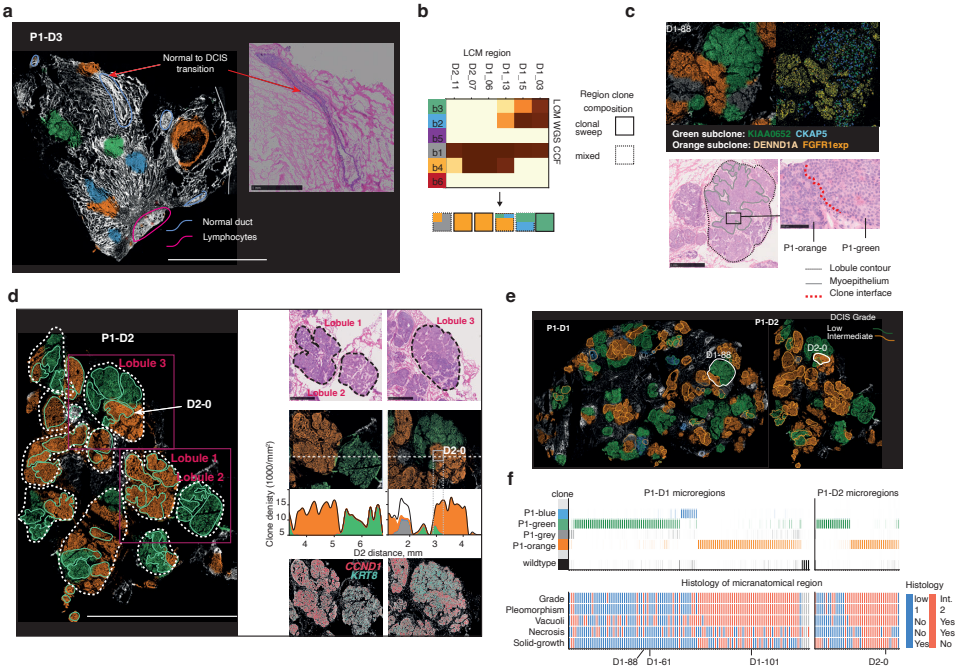
Variable	P2-TN1	P2-TN2	P2-LN	P1-ER1	P1-ER2	P1-D1	P1-D2	P1-D3
Manuscript ID	P2	P2	P2	P1	P1	P1	P1	P1
Case	PD14780	PD14780	PD14780	PD9694	PD9694	PD9694	PD9694	PD9694
Sub-sample	a	d	e	a	c	d	l	m
Type	Primary invasive tumour	Primary invasive tumour	Lymph node with metastasis	Primary invasive tumour	Primary invasive tumour	DCIS	DCIS	DCIS
Path Review (Frozen H&E Summary)	Admixed carcinoma, stroma and lymphocytes scattered throughout but in differing proportions.	Admixed carcinoma and stroma, lymphoid island and necrotic areas	Lymph node replaced by carcinoma, cannot see capsule. Dense islands of tumour, may be in sinuses or discrete deposits.	Mainly invasive tumour (NST) with areas of DCIS	Mainly invasive tumour (NST) with areas of DCIS	DCIS and cancerisation of lobules (cribiform features, necrosis, combination of intermediate grade and intermediate to high grade)	DCIS and cancerisation of lobules (cribiform features, necrosis, combination of intermediate grade and intermediate to high grade)	DCIS and cancerisation of lobules, several normal lobules and a duct transforming from DCIS to normal along length
Tissue Size (mm^2)	38	68.5	44	62.36	51.3	132	34	58
Nuclei Count	224826	261305	394895	319925	298682	642256	173508	104523



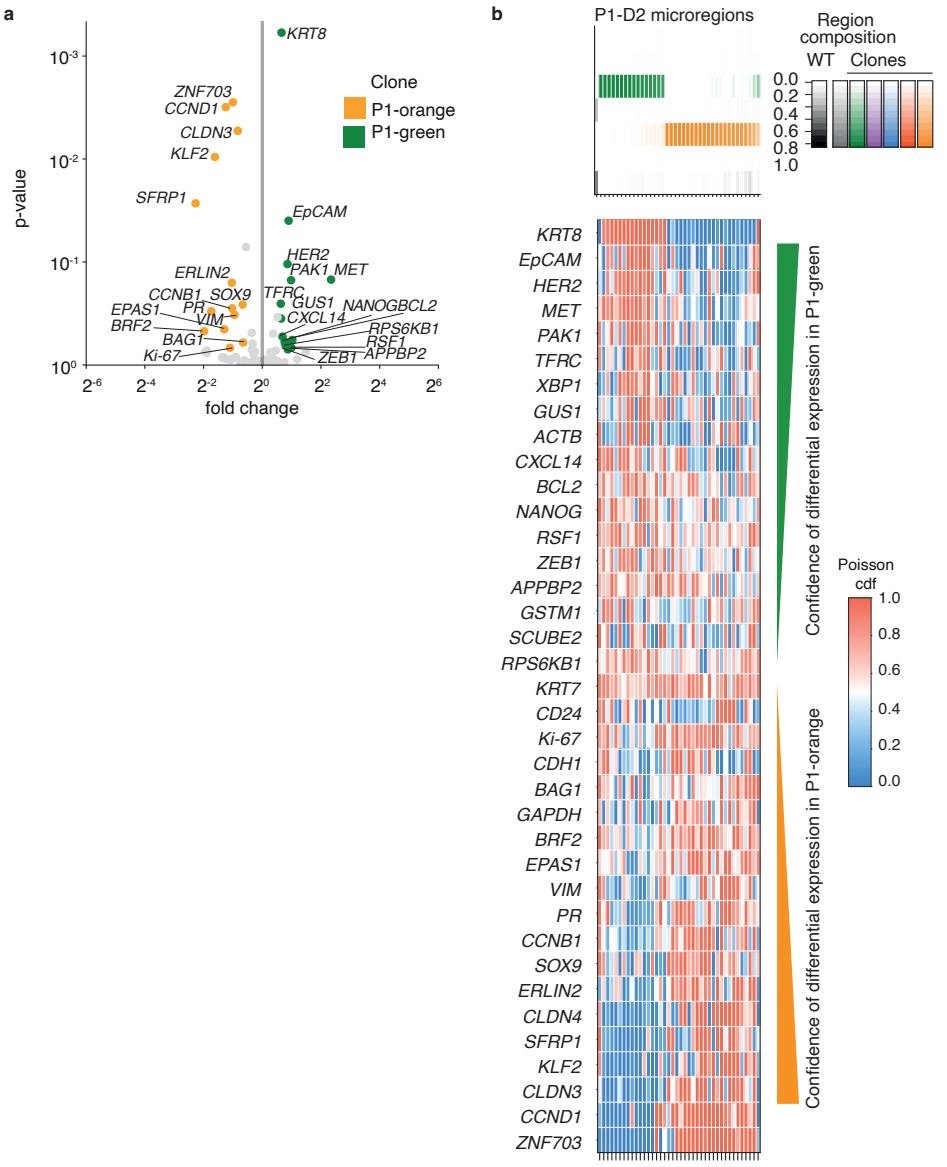
Appendix Figure D.1: Phenotype characterisation of histo-genomic states in sample P1 PBCs. **a**, Broadly annotated H&E tissue sections of the P1-ER1 and P1-ER2 primary breast cancers. **b**, Microregions selected for detailed analysis overlaid on BaSISS maps (regions relate to heatmaps in Fig. 3a; numbers relate to histological annotations in Supplementary Table 2). c, Comparison of the cancer cell fraction (CCF) of 9 regions of P1-ER1/P1-ER2 determined through both BaSISS (top) and laser capture microdissection (LCM) whole genome sequencing (WGS) (bottom). **d**, Snapshots of immunohistochemistry (IHC) staining in serial fresh frozen tissue cryosections from P1-ER2. Selected regions with confirmed clone compositions (by LCM-WGS) are presented. SMMHC/P63 antibody stains myoepithelial cells red, PTEN protein and the progesterone receptor (PR) stain brown. % reports proportion of positive nuclei stained, n reports number of nuclei in region assessed by QuPath digital software. Row 1–3 scale bars = 250 μm. Row 4 scale bar = 50 μm. **e**, Violin plots depict clone specific Ki67 IHC staining rate posterior density of the generalised linear mixed model (glmm) with region specific random effect. Significant comparisons were controlled for FDR using the Benjamini-Hochberg procedure. Analysis was limited to the 11 regions with confirmed clone compositions by WGS due to variation between IHC and BaSISS sections in z-stack morphology (relates to the bottom panel on Figure 2.2). **f**, Violin plots depict clone specific gene expression contribution posterior density of the glmm with region specific random effect. A total of 36 regions of P1-ER2 with a dominant clone fraction > 0.7 were analysed. Significant comparisons were controlled for FDR using the Benjamini-Hochberg procedure. DCIS - Ductal carcinoma in situ.



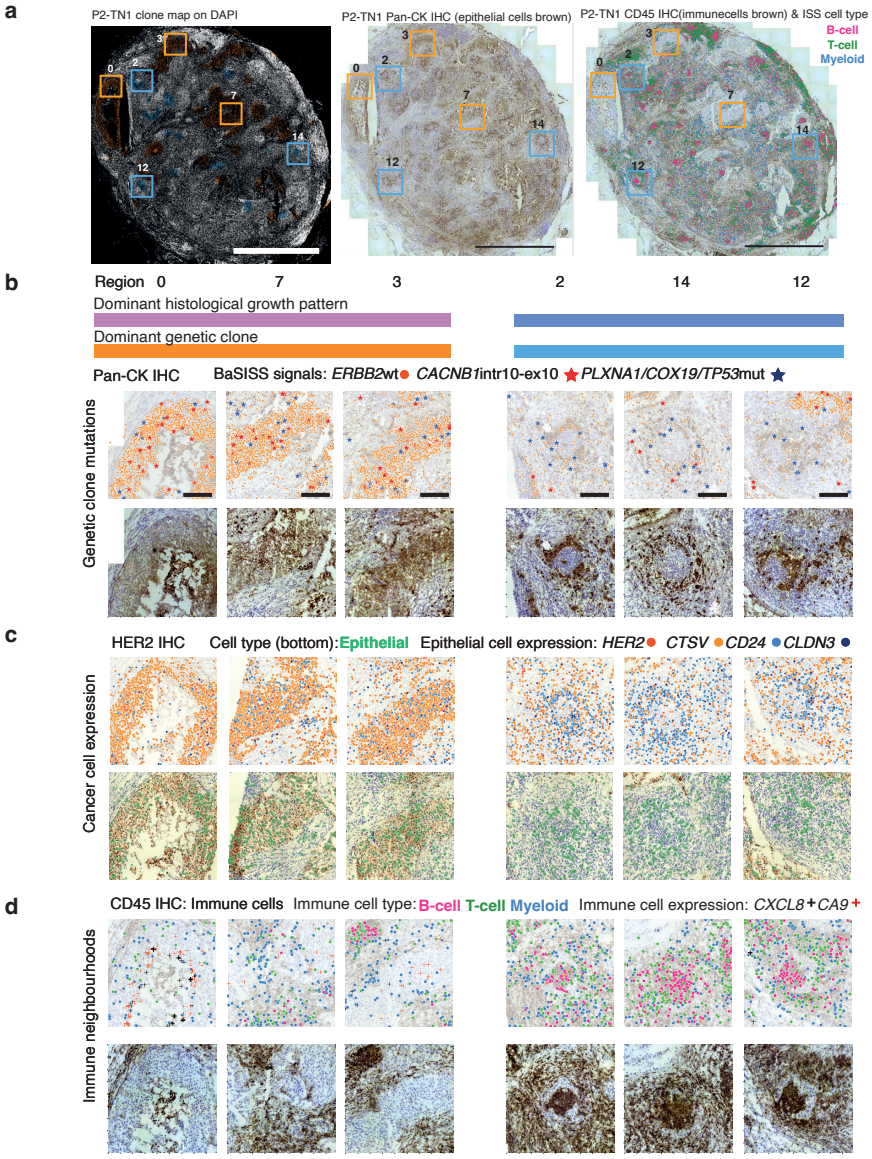
Appendix Figure D.2: Ecosystem characterisation in P2-TN1. **a**, H&E stained sections of the two primary breast cancers from case P2. **b**, Microregions selected for detailed analysis overlaid on BaSISS maps (regions relate to heatmaps in top Figure 4.8; numbers relate to histological annotations in Supplementary Table 2). Microregions were not defined for P2-TN2 as a single clone was targeted and detected. **c**, Cell type contribution posterior density of the generalised linear mixed models (glmm) model with region specific random effect. Significant comparisons were controlled for FDR using the Benjamini-Hochberg procedure. 19 clone territories (with dominant clone fraction > 0.1) were analysed. Fibroblasts and perivascular-like cells (PVL) could not be differentiated within this experiment and are reported as 'fibroblasts'. **d**, Volcano plot of epithelial expression of the 91 oncology ISS panel genes in TN1 invasive regions. Significance was adjusted for multiple testing using BH procedure, only genes with FDR < 0.1 and fold change > 1.5 in both ways are coloured/labelled. DCIS = Ductal carcinoma in situ.



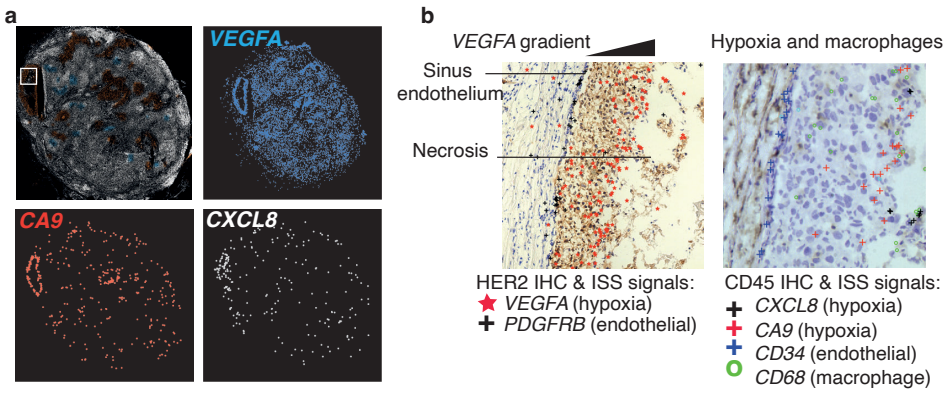
Appendix Figure D.3: DCIS clone specific histologies. **a**, BaSISS clone map of P1-D3, a sample that contains Ductal carcinoma in Situ (DCIS), stroma and normal glandular regions. The most prevalent genetic clone colour is projected as a coloured field on DAPI images (reported if cancer cell fraction > 25% and inferred local cell density > 300 cells/mm2). Scale bar = 5 mm. Inlaid, H&E stained image (from a serial tissue section) details the histological transition from normal to DCIS morphology, consistent with the clone field transition in the BaSISS map (scale bar = 1 mm). **b**, Heatmap of CCF derived from LCM-WGS of six regions of P1-D1/P1-D2 with cartoon of predicted clone composition indicating inference of monoclonal and polyclonal growth patterns. **c**, Example of a clone interface within a single sub-lobular space in P1-D1. Clone fields (top left); spatial BaSISS mutation signals (top right); characteristic histological features on H&E (bottom left) with zoom image of clone interface (scale bar = 100 μm) (bottom right). **d**, Histological, genetic and transcriptional features of three lobules (identified on the clone map of P1-D2; left, scale bar = 5 mm) are shown: H&E staining (top) scale bar = 1 mm; BaSISS clone fields projected on DAPI with frequency plots of the local, mean cancer (coloured areas) and non-cancer (white) corresponding to horizontal dashed line (middle); and ISS gene expression signals reporting *CCND1* and *KRT8* that exhibit clone specific spatial patterns. **e**, Clone maps of P1-D1/P1-D2 (as presented in Figure 4.9) but microregions are coloured according to histological grade. **f**, Histopathological annotations for each microregion presented alongside the same clone composition heatmap as shown in Figure 4.9.



Appendix Figure D.4: Distinct transcriptional profiles of two DCIS clones. **a**, Volcano plot of epithelial expression of the 91 oncology ISS panel genes in P1-D2. Significance was adjusted for multiple testing using BH procedure, only genes with FDR < 0.1 and fold change > 1.5 in both ways are coloured/labelled. The coloured genes are included in the by region plot in **b**. **b**, Heatmap of gene expression data within each of the 41 sampled regions in P1-D2, presented alongside the relevant clone composition regions (top) as per Figure 4.9. ISS counts in each regions are transformed by applying Poisson cdf with λ = mean (P1-green expression, P1-orange expression) $\times$ nuclei count in each region, thus divergence from 0.5 reflects deviation from the global mean expression. Only genes with FDR < 0.1 are presented and ordered by the confidence of differential.



Appendix Figure D.5: Highly recurrent clone specific ecosystems in a metastatic lymph node. **a**, P2-LN1 sample (left) DAPI image with BaSISS subclone fields (as shown in Figure 4.11) and coloured squares mark regions depicted in **b,c,d**; (middle) pan-cytokeratin IHC stained (epithelial cells appear brown); (right) CD45 antibody (immune cells appear brown) with ISS immune panel derived cell types projected as coloured dots. **b**—**d**, Snapshots of example regions dominated by P2-blue or P2-orange clones, as indicated in **a**. In each case signals (dots) from selected targets in BaSISS **b**, ISS oncology **c** or ISS immune panels **d** are presented overlaid on sections stained by IHC following the BaSISS/ISS experiment. In the bottom row of **c** and top row of **d** inferred epithelial and immune cell types are presented. In top rows of **c** and **d**, 80% transparency is applied to the underlying IHC image to aid visualisation of overlaid dots.



Appendix Figure D.6: Hypoxic signature in a metastatic lymph node. **a**, Spatial patterns of three hypoxia related genes are projected on the entire P2-LN1 tissue section. **b**, Spatial patterns of *PDGFRB*, *CD34*, *CD68* and hypoxia related ISS signals overlaid on HER2 (left) and CD45 IHC stained sections(right) correspond to region of white square on top left clone field image in **a**.

Bibliography

't Veer, Laura J van, Hongyue Dai, Marc J van de Vijver, Yudong D He, Augustinus A M Hart, Mao Mao, Hans L Peterse, Karin van der Kooy, Matthew J Marton, Anke T Witteveen, George J Schreiber, Ron M Kerkhoven, Chris Roberts, Peter S Linsley, René Bernards and Stephen H Friend (Jan. 2002). 'Gene expression profiling predicts clinical outcome of breast cancer'. en. In: *Nature* 415.6871, pp. 530–536.

Abadi, Martín, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu and Xiaoqiang Zheng (May 2016). 'TensorFlow: A system for large-scale machine learning'. In: arXiv: 1605.08695 [cs.DC].

Abby, Emilie, Stefan C Dentro, Michael W J Hall, Joanna C Fowler, Swee Hoe Ong, Roshan Sood, Christian W Siebel, Moritz Gerstung, Benjamin A Hall and Philip H Jones (June 2021). 'Notch1 mutation drives clonal expansion in normal esophageal epithelium but impairs tumor growth'. en.

Aceto, Nicola, Aditya Bardia, David T Miyamoto, Maria C Donaldson, Ben S Wittner, Joel A Spencer, Min Yu, Adam Pely, Amanda Engstrom, Huili Zhu, Brian W Brannigan, Ravi Kapur, Shannon L Stott, Toshi Shioda, Sridhar Ramaswamy, David T Ting, Charles P Lin, Mehmet Toner, Daniel A Haber and Shyamala Maheswaran (Aug. 2014). 'Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis'. en. In: *Cell* 158.5, pp. 1110–1122.

Adams, Daniel L, Stuart S Martin, R Katherine Alpaugh, Monica Charpentier, Susan Tsai, Raymond C Bergan, Irene M Ogden, William Catalona, Saranya Chumsri, Cha-Mei Tang and Massimo Cristofanilli (Mar. 2014). 'Circulating giant macrophages as a potential biomarker of solid tumors'. en. In: *Proc. Natl. Acad. Sci. U. S. A.* 111.9, pp. 3514–3519.

Ahlers, Jannis, Daniel Althviz Moré, Oren Amsalem, Ashley Anderson, Grzegorz Bokota, Peter Boone, Jordão Bragantini, Genevieve Buckley, Alister Burt, Matthias Bussonnier, Ahmet Can Solak, Clément Caporal, Draga Doncila Pop, Kira Evans, Jeremy Freeman, Lorenzo Gaifas, Christoph Gohlke, Kabilar Gunalan, Hagai Har-Gil, Mark Harfouche, Kyle I S Harrington, Volker Hilsenstein, Katherine Hutchings, Talley Lambert, Jessy Lauer, Gregor Lichtner, Ziyang Liu, Lucy Liu, Alan Lowe, Luca Marconato, Sean Martin, Abigail McGovern, Lukasz Migas, Nadalyn Miller, Hector Muñoz, Jan-Hendrik Müller, Christopher Nauroth-Kreß, Juan Nunez-Iglesias, Constantin Pape, Kim Pevey, Gonzalo Peña-Castellanos, Andrea Pierré, Jaime Rodríguez-Guerra, David

Ross, Loic Royer, Craig T Russell, Gabriel Selzer, Paul Smith, Peter Sobolewski, Konstantin Sofiiuk, Nicholas Sofroniew, David Stansby, Andrew Sweet, Wouter-Michiel Vierdag, Pam Wadhwa, Melissa Weber Mendonça, Jonas Windhager, Philip Winston and Kevin Yamauchi (July 2023). *napari: a multi-dimensional image viewer for Python*.

Alexandrov, Ludmil B, Jaegil Kim, Nicholas J Haradhvala, Mi Ni Huang, Alvin Wei Tian Ng, Yang Wu, Arnoud Boot, Kyle R Covington, Dmitry A Gordenin, Erik N Bergstrom, S M Ashiqul Islam, Nuria Lopez-Bigas, Leszek J Klimczak, John R McPherson, Sandro Morganella, Radhakrishnan Sabarinathan, David A Wheeler, Ville Mustonen, PCAWG Mutational Signatures Working Group, Gad Getz, Steven G Rozen, Michael R Stratton and PCAWG Consortium (Feb. 2020). 'The repertoire of mutational signatures in human cancer'. en. In: *Nature* 578.7793, pp. 94–101.

Ali, H Raza, Hartland W Jackson, Vito R T Zanotelli, Esther Danenberg, Jana R Fischer, Helen Bardwell, Elena Provenzano, CRUK IMAXT Grand Challenge Team, Oscar M Rueda, Suet-Feung Chin, Samuel Aparicio, Carlos Caldas and Bernd Bodenmiller (Feb. 2020). 'Imaging mass cytometry and multiplatform genomics define the phenogenomic landscape of breast cancer'. en. In: *Nat Cancer* 1.2, pp. 163–175.

Allan, Chris, Jean-Marie Burel, Josh Moore, Colin Blackburn, Melissa Linkert, Scott Loynton, Donald Macdonald, William J Moore, Carlos Neves, Andrew Patterson, Michael Porter, Aleksandra Tarkowska, Brian Loranger, Jerome Avondo, Ingvar Lagerstedt, Luca Lianas, Simone Leo, Katherine Hands, Ron T Hay, Ardan Patwardhan, Christoph Best, Gerard J Kleywegt, Gianluigi Zanetti and Jason R Swedlow (Feb. 2012). 'OMERO: flexible, model-driven data management for experimental biology'. en. In: *Nat. Methods* 9.3, pp. 245–253.

Alon, Shahar, Daniel R Goodwin, Anubhav Sinha, Asmamaw T Wassie, Fei Chen, Evan R Daugharthy, Yosuke Bando, Atsushi Kajita, Andrew G Xue, Karl Marrett, Robert Prior, Yi Cui, Andrew C Payne, Chun-Chen Yao, Ho-Jun Suk, Ru Wang, Chih-Chieh Jay Yu, Paul Tillberg, Paul Reginato, Nikita Pak, Songlei Liu, Sukanya Punthambaker, Eswar P R Iyer, Richie E Kohman, Jeremy A Miller, Ed S Lein, Ana Lako, Nicole Cullen, Scott Rodig, Karla Helvie, Daniel L Abravanel, Nikhil Wagle, Bruce E Johnson, Johanna Klughammer, Michal Slyper, Julia Waldman, Judit Jané-Valbuena, Orit Rozenblatt-Rosen, Aviv Regev, IMAXT Consortium, George M Church, Adam H Marblestone and Edward S Boyden (Jan. 2021). 'Expansion sequencing: Spatially precise in situ transcriptomics in intact biological systems'. en. In: *Science* 371.6528.

Alonso-Curbelo, Direna, Yu-Jui Ho, Cassandra Burdziak, Jesper L V Maag, John P Morris 4th, Rohit Chandwani, Hsuan-An Chen, Kaloyan M Tsanov, Francisco M Barriga, Wei Luan, Nilgun Tasdemir, Geulah Livshits, Elham Azizi, Jaeyoung Chun, John E Wilkinson, Linas Mazutis, Steven D Leach, Richard Koche, Dana Pe'er and Scott W Lowe (Feb. 2021). 'A gene-environment-induced epigenetic program initiates tumorigenesis'. en. In: *Nature* 590.7847, pp. 642–648.

Andersson, Alma, Ludvig Larsson, Linnea Stenbeck, Fredrik Salmén, Anna Ehinger, Sunny Z Wu, Ghamdan Al-Eryani, Daniel Roden, Alex Swarbrick, Åke Borg, Jonas Frisén, Camilla Engblom and Joakim Lundeberg (Oct. 2021). 'Spatial deconvolution of

HER2-positive breast cancer delineates tumor-associated cell type interactions'. en. In: *Nat. Commun.* 12.1, p. 6012.

Andor, Noemi, Trevor A Graham, Marnix Jansen, Li C Xia, C Athena Aktipis, Claudia Petritsch, Hanlee P Ji and Carlo C Maley (Jan. 2016). 'Pan-cancer analysis of the extent and consequences of intratumor heterogeneity'. en. In: *Nat. Med.* 22.1, pp. 105–113.

Angelo, Michael, Sean C Bendall, Rachel Finck, Matthew B Hale, Chuck Hitzman, Alexander D Borowsky, Richard M Levenson, John B Lowe, Scot D Liu, Shuchun Zhao, Yasodha Natkunam and Garry P Nolan (Apr. 2014). 'Multiplexed ion beam imaging of human breast tumors'. en. In: *Nat. Med.* 20.4, pp. 436–442.

Aran, Dvir, Marina Sirota and Atul J Butte (Dec. 2015). 'Systematic pan-cancer analysis of tumour purity'. en. In: *Nat. Commun.* 6, p. 8971.

Arganda-Carreras, I, C Sorzano, J Kybic and C Ortiz-de-Solórzano (2008). 'bUnwarpJ : Consistent and Elastic Registration in ImageJ . Methods and Applications'. In: *undefined.*

Argelaguet, Ricard, Britta Velten, Damien Arnol, Sascha Dietrich, Thorsten Zenz, John C Marioni, Florian Buettner, Wolfgang Huber and Oliver Stegle (June 2018). 'Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets'. en. In: *Mol. Syst. Biol.* 14.6, e8124.

Arnol, Damien, Denis Schapiro, Bernd Bodenmiller, Julio Saez-Rodriguez and Oliver Stegle (Oct. 2019). 'Modeling Cell-Cell Interactions from Spatial Molecular Data with Spatial Variance Component Analysis'. en. In: *Cell Rep.* 29.1, 202–211.e6.

Ashuach, Tal, Mariano I Gabitto, Rohan V Koodli, Giuseppe-Antonio Saldi, Michael I Jordan and Nir Yosef (Aug. 2023). 'MultiVI: deep generative model for the integration of multimodal data'. en. In: *Nat. Methods* 20.8, pp. 1222–1231.

Azimi, Farhad, Richard A Scolyer, Pavlina Rumcheva, Marc Moncrieff, Rajmohan Murali, Stanley W McCarthy, Robyn P Saw and John F Thompson (July 2012). 'Tumor-infiltrating lymphocyte grade is an independent predictor of sentinel lymph node status and survival in patients with cutaneous melanoma'. en. In: *J. Clin. Oncol.* 30.21, pp. 2678–2683.

Azizi, Elham, Ambrose J Carr, George Plitas, Andrew E Cornish, Catherine Konopacki, Sandhya Prabhakaran, Juozas Nainys, Kenmin Wu, Vaidotas Kiseliovas, Manu Setty, Kristy Choi, Rachel M Fromme, Phuong Dao, Peter T McKenney, Ruby C Wasti, Krishna Kadaveru, Linas Mazutis, Alexander Y Rudensky and Dana Pe'er (Aug. 2018). 'Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment'. en. In: *Cell* 174.5, 1293–1308.e36.

Baker, Ann-Marie, Weini Huang, Xiao-Ming Mindy Wang, Marnix Jansen, Xiao-Jun Ma, Jeffrey Kim, Courtney M Anderson, Xingyong Wu, Liuliu Pan, Nan Su, Yuling Luo, Enric Domingo, Timon Heide, Andrea Sottoriva, Annabelle Lewis, Andrew D Beggs, Nicholas A Wright, Manuel Rodriguez-Justo, Emily Park, Ian Tomlinson and Trevor A Graham (Dec. 2017). 'Robust RNA-based in situ mutation detection delineates colorectal cancer subclonal evolution'. en. In: *Nat. Commun.* 8.1, p. 1998.

Balachandran, Vinod P, Marta Łuksza, Julia N Zhao, Vladimir Makarov, John Alec Moral, Romain Remark, Brian Herbst, Gokce Askan, Umesh Bhanot, Yasin Senbabaoglu, Daniel K Wells, Charles Ian Ormsby Cary, Olivera Grbovic-Huezo, Marc Attiyeh, Benjamin Medina, Jennifer Zhang, Jennifer Loo, Joseph Saglimbeni, Mohsen Abu-Akeel, Roberta Zappasodi, Nadeem Riaz, Martin Smoragiewicz, Z Larkin Kelley, Olca Basturk, Australian Pancreatic Cancer Genome Initiative, Garvan Institute of Medical Research, Prince of Wales Hospital, Royal North Shore Hospital, University of Glasgow, St Vincent's Hospital, QIMR Berghofer Medical Research Institute, University of Melbourne, Centre for Cancer Research, University of Queensland, Institute for Molecular Bioscience, Bankstown Hospital, Liverpool Hospital, Royal Prince Alfred Hospital, Chris O'Brien Lifehouse, Westmead Hospital, Fremantle Hospital, St John of God Healthcare, Royal Adelaide Hospital, Flinders Medical Centre, Envoi Pathology, Princess Alexandria Hospital, Austin Hospital, Johns Hopkins Medical Institutes, ARC-Net Centre for Applied Research on Cancer, Mithat Gönen, Arnold J Levine, Peter J Allen, Douglas T Fearon, Miriam Merad, Sacha Gnjatic, Christine A Iacobuzio-Donahue, Jedd D Wolchok, Ronald P DeMatteo, Timothy A Chan, Benjamin D Greenbaum, Taha Merghoub and Steven D Leach (Nov. 2017). 'Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer'. en. In: *Nature* 551.7681, pp. 512–516.

Baldominos, Pilar, Alex Barbera-Mourelle, Olga Barreiro, Yu Huang, Andrew Wight, Jae-Won Cho, Xi Zhao, Guillem Estivill, Isam Adam, Xavier Sanchez, Shannon McCarthy, Julien Schaller, Zara Khan, Albert Ruzo, Ricardo Pastorello, Edward T Richardson, Deborah Dillon, Paula Montero-Llopis, Romualdo Barroso-Sousa, Juliet Forman, Sachet A Shukla, Sara M Tolaney, Elizabeth A Mittendorf, Ulrich H von Andrian, Kai W Wucherpfennig, Martin Hemberg and Judith Agudo (May 2022). 'Quiescent cancer cells resist T cell attack by forming an immunosuppressive niche'. en. In: *Cell* 185.10, 1694–1708.e19.

Banerji, Shantanu, Kristian Cibulskis, Claudia Rangel-Escareno, Kristin K Brown, Scott L Carter, Abbie M Frederick, Michael S Lawrence, Andrey Y Sivachenko, Carrie Sougnez, Lihua Zou, Maria L Cortes, Juan C Fernandez-Lopez, Shouyong Peng, Kristin G Ardlie, Daniel Auclair, Veronica Bautista-Piña, Fujiko Duke, Joshua Francis, Joonil Jung, Antonio Maffuz-Aziz, Robert C Onofrio, Melissa Parkin, Nam H Pho, Valeria Quintanar-Jurado, Alex H Ramos, Rosa Rebollar-Vega, Sergio Rodriguez-Cuevas, Sandra L Romero-Cordoba, Steven E Schumacher, Nicolas Stransky, Kristin M Thompson, Laura Uribe-Figueroa, Jose Baselga, Rameen Beroukhim, Kornelia Polyak, Dennis C Sgroi, Andrea L Richardson, Gerardo Jimenez-Sanchez, Eric S Lander, Stacey B Gabriel, Levi A Garraway, Todd R Golub, Jorge Melendez-Zajgla, Alex Toker, Gad Getz, Alfredo Hidalgo-Miranda and Matthew Meyerson (June 2012). 'Sequence analysis of mutations and translocations across breast cancer subtypes'. en. In: *Nature* 486.7403, pp. 405–409.

Bankhead, Peter, Maurice B Loughrey, José A Fernández, Yvonne Dombrowski, Darragh G McArt, Philip D Dunne, Stephen McQuaid, Ronan T Gray, Liam J Murray, Helen G

Coleman, Jacqueline A James, Manuel Salto-Tellez and Peter W Hamilton (Dec. 2017). 'QuPath: Open source software for digital pathology image analysis'. en. In: *Sci. Rep.* 7.1, p. 16878.

Bao, Li, Zhaoyang Qian, Maria B Lyng, Ling Wang, Yuan Yu, Ting Wang, Xiuqing Zhang, Huanming Yang, Nils Brünner, Jun Wang and Henrik J Ditzel (June 2018). 'Coexisting genomic aberrations associated with lymph node metastasis in breast cancer'. en. In: *J. Clin. Invest.* 128.6, pp. 2310–2324.

Barkley, Dalia, Reuben Moncada, Maayan Pour, Deborah A Liberman, Ian Dryg, Gregor Werba, Wei Wang, Maayan Baron, Anjali Rao, Bo Xia, Gustavo S França, Alejandro Weil, Deborah F Delair, Cristina Hajdu, Amanda W Lund, Iman Osman and Itai Yanai (Aug. 2022). 'Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment'. en. In: *Nat. Genet.* 54.8, pp. 1192–1201.

Barry, Peter, Alexandra Vatsiou, Inmaculada Spiteri, Daniel Nichol, George D Cresswell, Ahmet Acar, Nicholas Trahearn, Sarah Hrebien, Isaac Garcia-Murillas, Kate Chkhaidze, Luca Ermini, Ian Said Huntingford, Hannah Cottom, Lila Zabaglo, Konrad Koelble, Saira Khalique, Jennifer E Rusby, Francesca Muscara, Mitch Dowsett, Carlo C Maley, Rachael Natrajan, Yinyin Yuan, Gaia Schiavon, Nicholas Turner and Andrea Sottoriva (Oct. 2018). 'The Spatiotemporal Evolution of Lymph Node Spread in Early Breast Cancer'. en. In: *Clin. Cancer Res.* 24.19, pp. 4763–4770.

Baslan, Timour, John P Morris 4th, Zhen Zhao, Jose Reyes, Yu-Jui Ho, Kaloyan M Tsanov, Jonathan Bermeo, Sha Tian, Sean Zhang, Gokce Askan, Aslihan Yavas, Nicolas Lecomte, Amanda Erakky, Anna M Varghese, Amy Zhang, Jude Kendall, Elena Ghiban, Lubomir Chorbadjiev, Jie Wu, Nevenka Dimitrova, Kalyani Chadalavada, Gouri J Nanjangud, Chaitanya Bandlamudi, Yixiao Gong, Mark T A Donoghue, Nicholas D Socci, Alex Krasnitz, Faiyaz Notta, Steve D Leach, Christine A Iacobuzio-Donahue and Scott W Lowe (Aug. 2022). 'Ordered and deterministic cancer genome evolution after p53 loss'. en. In: *Nature* 608.7924, pp. 795–802.

Baysoy, Alev, Zhiliang Bai, Rahul Satija and Rong Fan (June 2023). 'The technological landscape and applications of single-cell multi-omics'. en. In: *Nat. Rev. Mol. Cell Biol.*, pp. 1–19.

Bergholtz, Helga, Jodi M Carter, Alessandra Cesano, Maggie Chon U Cheang, Sarah E Church, Prajan Divakar, Christopher A Fuhrman, Shom Goel, Jingjing Gong, Jennifer L Guerriero, Margaret L Hoang, E Shelley Hwang, Hellen Kuasne, Jinho Lee, Yan Liang, Elizabeth A Mittendorf, Jessica Perez, Aleix Prat, Lajos Pusztai, Jason W Reeves, Yasser Riazalhosseini, Jennifer K Richer, Özgür Sahin, Hiromi Sato, Ilana Schlam, Therese Sørlie, Daniel G Stover, Sandra M Swain, Alexander Swarbrick, E Aubrey Thompson, Sara M Tolaney, Sarah E Warren and On Behalf Of The GeoMx Breast Cancer Consortium (Sept. 2021). 'Best Practices for Spatial Profiling for Breast Cancer Research with the GeoMx® Digital Spatial Profiler'. en. In: *Cancers* 13.17, p. 4456.

Berglund, Emelie, Jonas Maaskola, Niklas Schultz, Stefanie Friedrich, Maja Marklund, Joseph Bergenstråhle, Firas Tarish, Anna Tanoglidi, Sanja Vickovic, Ludvig Larsson, Fredrik Salmén, Christoph Ogris, Karolina Wallenborg, Jens Lagergren, Patrik Ståhl,

Erik Sonnhammer, Thomas Helleday and Joakim Lundeberg (June 2018). 'Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity'. en. In: *Nat. Commun.* 9.1, p. 2419.

Bingham, Eli, Jonathan P Chen, Martin Jankowiak, Fritz Obermeyer, Neeraj Pradhan, Theofanis Karaletsos, Rohit Singh, Paul Szerlip, Paul Horsfall and Noah D Goodman (Oct. 2018). 'Pyro: Deep Universal Probabilistic Programming'. In: *J. Mach. Learn. Res.* 28, pp. 1–6. arXiv: 1810.09538 [cs.LG].

Biswas, Dhruva, Nicolai J Birkbak, Rachel Rosenthal, Crispin T Hiley, Emilia L Lim, Krisztian Papp, Stefan Boeing, Marcin Krzystanek, Dijana Djureinovic, Linnea La Fleur, Maria Greco, Balázs Döme, János Fillinger, Hans Brunnström, Yin Wu, David A Moore, Marcin Skrzypski, Christopher Abbosh, Kevin Litchfield, Maise Al Bakir, Thomas B K Watkins, Selvaraju Veeriah, Gareth A Wilson, Mariam Jamal-Hanjani, Judit Moldvay, Johan Botling, Arul M Chinnaiyan, Patrick Micke, Allan Hackshaw, Jiri Bartek, Istvan Csabai, Zoltan Szallasi, Javier Herrero, Nicholas McGranahan, Charles Swanton and TRACERx Consortium (Oct. 2019). 'A clonal expression biomarker associates with lung cancer mortality'. en. In: *Nat. Med.* 25.10, pp. 1540–1548.

Blei, David M, Alp Kucukelbir and Jon D McAuliffe (Apr. 2017). 'Variational Inference: A Review for Statisticians'. In: *J. Am. Stat. Assoc.* 518, pp. 859–877. arXiv: 1601.00670 [stat.CO].

Bolli, Niccolo, Hervé Avet-Loiseau, David C Wedge, Peter Van Loo, Ludmil B Alexandrov, Inigo Martincorena, Kevin J Dawson, Francesco Iorio, Serena Nik-Zainal, Graham R Bignell, Jonathan W Hinton, Yilong Li, Jose M C Tubio, Stuart McLaren, Sarah O' Meara, Adam P Butler, Jon W Teague, Laura Mudie, Elizabeth Anderson, Naim Rashid, Yu-Tzu Tai, Masood A Shammas, Adam S Sperling, Mariateresa Fulciniti, Paul G Richardson, Giovanni Parmigiani, Florence Magrangeas, Stephane Minvielle, Philippe Moreau, Michel Attal, Thierry Facon, P Andrew Futreal, Kenneth C Anderson, Peter J Campbell and Nikhil C Munshi (2014). 'Heterogeneity of genomic evolution and mutational profiles in multiple myeloma'. en. In: *Nat. Commun.* 5, p. 2997.

Borst, Jannie, Tomasz Ahrends, Nikolina Bąbała, Cornelis J M Melief and Wolfgang Kastenmüller (Oct. 2018). 'CD4+ T cell help in cancer immunology and immunotherapy'. en. In: *Nat. Rev. Immunol.* 18.10, pp. 635–647.

Bradbury, James, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne et al. (2018). 'JAX: composable transformations of Python+ NumPy programs'. In.

BRADSKI and G (2000). 'The OpenCV library'. In: *Dr Dobb's J. Software Tools* 25, pp. 120–125.

Brastianos, Priscilla K, Scott L Carter, Sandro Santagata, Daniel P Cahill, Amaro Taylor-Weiner, Robert T Jones, Eliezer M Van Allen, Michael S Lawrence, Peleg M Horowitz, Kristian Cibulskis, Keith L Ligon, Josep Tabernero, Joan Seoane, Elena Martinez-Saez, William T Curry, Ian F Dunn, Sun Ha Paek, Sung-Hye Park, Aaron McKenna, Aaron Chevalier, Mara Rosenberg, Frederick G Barker 2nd, Corey M Gill, Paul Van Hummelen, Aaron R Thorner, Bruce E Johnson, Mai P Hoang, Toni K Choueiri,

Sabina Signoretti, Carrie Sougnez, Michael S Rabin, Nancy U Lin, Eric P Winer, Anat Stemmer-Rachamimov, Matthew Meyerson, Levi Garraway, Stacey Gabriel, Eric S Lander, Rameen Beroukhim, Tracy T Batchelor, Jose Baselga, David N Louis, Gad Getz and William C Hahn (Nov. 2015). 'Genomic Characterization of Brain Metastases Reveals Branched Evolution and Potential Therapeutic Targets'. en. In: *Cancer Discov.* 5.11, pp. 1164–1177.

Brown, David, Dominiek Smeets, Borbála Székely, Denis Larsimont, A Marcell Szász, Pierre-Yves Adnet, Françoise Rothé, Ghizlane Rouas, Zsófia I Nagy, Zsófia Faragó, Anna-Mária Tőkés, Magdolna Dank, Gyöngyvér Szentmártoni, Nóra Udvarhelyi, Gabriele Zoppoli, Lajos Pusztai, Martine Piccart, Janina Kulka, Diether Lambrechts, Christos Sotiriou and Christine Desmedt (Apr. 2017). 'Phylogenetic analysis of metastatic progression in breast cancer using somatic mutations and copy number aberrations'. en. In: *Nat. Commun.* 8, p. 14944.

Bruin, Elza C de, Nicholas McGranahan, Richard Mitter, Max Salm, David C Wedge, Lucy Yates, Mariam Jamal-Hanjani, Seema Shafi, Nirupa Murugaesu, Andrew J Rowan, Eva Grönroos, Madiha A Muhammad, Stuart Horswell, Marco Gerlinger, Ignacio Varela, David Jones, John Marshall, Thierry Voet, Peter Van Loo, Doris M Rassl, Robert C Rintoul, Sam M Janes, Siow-Ming Lee, Martin Forster, Tanya Ahmad, David Lawrence, Mary Falzon, Arrigo Capitanio, Timothy T Harkins, Clarence C Lee, Warren Tom, Enock Teefe, Shann-Ching Chen, Sharmin Begum, Adam Rabinowitz, Benjamin Phillimore, Bradley Spencer-Dene, Gordon Stamp, Zoltan Szallasi, Nik Matthews, Aengus Stewart, Peter Campbell and Charles Swanton (Oct. 2014). 'Spatial and temporal diversity in genomic instability processes defines lung cancer evolution'. en. In: *Science* 346.6206, pp. 251–256.

Cairns, J (May 1975). 'Mutation selection and the natural history of cancer'. en. In: *Nature* 255.5505, pp. 197–200.

Cairns, Rob A and Richard P Hill (Mar. 2004). 'Acute hypoxia enhances spontaneous lymph node metastasis in an orthotopic murine model of human cervical carcinoma'. en. In: *Cancer Res.* 64.6, pp. 2054–2061.

Cancer Genome Atlas Network (Oct. 2012). 'Comprehensive molecular portraits of human breast tumours'. en. In: *Nature* 490.7418, pp. 61–70.

Cancer Genome Atlas Research Network, John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Mills Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander and Joshua M Stuart (Oct. 2013). 'The Cancer Genome Atlas Pan-Cancer analysis project'. en. In: *Nat. Genet.* 45.10, pp. 1113–1120.

Casasent, Anna K, Aislyn Schalck, Ruli Gao, Emi Sei, Annalyssa Long, William Pangburn, Tod Casasent, Funda Meric-Bernstam, Mary E Edgerton and Nicholas E Navin (Jan. 2018). 'Multiclonal Invasion in Breast Tumors Identified by Topographic Single Cell Sequencing'. en. In: *Cell* 172.1-2, 205–217.e12.

Chang, H W, M Aoki, D Fruman, K R Auger, A Bellacosa, P N Tsichlis, L C Cantley, T M Roberts and P K Vogt (June 1997). 'Transformation of chicken cells by the gene encoding the catalytic subunit of PI 3-kinase'. en. In: *Science* 276.5320, pp. 1848–1850.

Chen, Ao, Sha Liao, Mengnan Cheng, Kailong Ma, Liang Wu, Yiwei Lai, Xiaojie Qiu, Jin Yang, Jiangshan Xu, Shijie Hao, Xin Wang, Huifang Lu, Xi Chen, Xing Liu, Xin Huang, Zhao Li, Yan Hong, Yujia Jiang, Jian Peng, Shuai Liu, Mengzhe Shen, Chuanyu Liu, Quanshui Li, Yue Yuan, Xiaoyu Wei, Huiwen Zheng, Weimin Feng, Zhifeng Wang, Yang Liu, Zhaohui Wang, Yunzhi Yang, Haitao Xiang, Lei Han, Baoming Qin, Pengcheng Guo, Guangyao Lai, Pura Muñoz-Cánoves, Patrick H Maxwell, Jean Paul Thiery, Qing-Feng Wu, Fuxiang Zhao, Bichao Chen, Mei Li, Xi Dai, Shuai Wang, Haoyan Kuang, Junhou Hui, Liqun Wang, Ji-Feng Fei, Ou Wang, Xiaofeng Wei, Haorong Lu, Bo Wang, Shiping Liu, Ying Gu, Ming Ni, Wenwei Zhang, Feng Mu, Ye Yin, Huanming Yang, Michael Lisby, Richard J Cornall, Jan Mulder, Mathias Uhlén, Miguel A Esteban, Yuxiang Li, Longqi Liu, Xun Xu and Jian Wang (May 2022). 'Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays'. en. In: *Cell* 185.10, 1777–1792.e21.

Chen, Benchao, Heng Li, Chao Liu, Xudong Xiang, Shuting Wang, Anhao Wu, Yan Shen and Gaofeng Li (Nov. 2020). 'Prognostic value of the common tumour-infiltrating lymphocyte subtypes for patients with non-small cell lung cancer: A meta-analysis'. en. In: *PLoS One* 15.11, e0242173.

Chen, Fei, Asmamaw T Wassie, Allison J Cote, Anubhav Sinha, Shahar Alon, Shoh Asano, Evan R Daugharthy, Jae-Byum Chang, Adam Marblestone, George M Church, Arjun Raj and Edward S Boyden (Aug. 2016). 'Nanoscale imaging of RNA with expansion microscopy'. en. In: *Nat. Methods* 13.8, pp. 679–684.

Chen, Kok Hao, Alistair N Boettiger, Jeffrey R Moffitt, Siyuan Wang and Xiaowei Zhuang (Apr. 2015). 'RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells'. en. In: *Science* 348.6233, aaa6090.

Chen, Xiaoyin, Yu-Chi Sun, George M Church, Je Hyuk Lee and Anthony M Zador (Feb. 2018). 'Efficient in situ barcode sequencing using padlock probe-based BaristaSeq'. en. In: *Nucleic Acids Res.* 46.4, e22.

Chérief-Abdellatif, Badr-Eddine (Oct. 2018). 'Consistency of ELBO maximization for model selection'. In: arXiv: 1810.11859 [math.ST].

Cheung, Kevin J, Veena Padmanaban, Vanesa Silvestri, Koen Schipper, Joshua D Cohen, Amanda N Fairchild, Michael A Gorin, James E Verdone, Kenneth J Pienta, Joel S Bader and Andrew J Ewald (Feb. 2016). 'Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters'. en. In: *Proc. Natl. Acad. Sci. U. S. A.* 113.7, E854–63.

Cho, Chun-Seok, Jingyue Xi, Yichen Si, Sung-Rye Park, Jer-En Hsu, Myungjin Kim, Goo Jun, Hyun Min Kang and Jun Hee Lee (June 2021). 'Microscopic examination of spatial transcriptome using Seq-Scope'. en. In: *Cell* 184.13, 3559–3572.e22.

Chung, Woosung, Hye Hyeon Eum, Hae-Ock Lee, Kyung-Min Lee, Han-Byoel Lee, Kyu-Tae Kim, Han Suk Ryu, Sangmin Kim, Jeong Eon Lee, Yeon Hee Park, Zhengyan Kan, Wonshik Han and Woong-Yang Park (May 2017). 'Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer'. en. In: *Nat. Commun.* 8, p. 15081.

Ciriello, Giovanni, Michael L Gatza, Andrew H Beck, Matthew D Wilkerson, Suhn K Rhie, Alessandro Pastore, Hailei Zhang, Michael McLellan, Christina Yau, Cyriac Kandoth, Reanne Bowlby, Hui Shen, Sikander Hayat, Robert Fieldhouse, Susan C Lester, Gary M K Tse, Rachel E Factor, Laura C Collins, Kimberly H Allison, Yunn-Yi Chen, Kristin Jensen, Nicole B Johnson, Steffi Oesterreich, Gordon B Mills, Andrew D Cherniack, Gordon Robertson, Christopher Benz, Chris Sander, Peter W Laird, Katherine A Hoadley, Tari A King, TCGA Research Network and Charles M Perou (Oct. 2015). 'Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer'. en. In: *Cell* 163.2, pp. 506–519.

Cleary, Allison S, Travis L Leonard, Shelley A Gestl and Edward J Gunther (Apr. 2014). 'Tumour cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers'. en. In: *Nature* 508.7494, pp. 113–117.

Codeluppi, Simone, Lars E Borm, Amit Zeisel, Gioele La Manno, Josina A van Lunteren, Camilla I Svensson and Sten Linnarsson (Nov. 2018). 'Spatial organization of the somatosensory cortex revealed by osmFISH'. en. In: *Nat. Methods* 15.11, pp. 932–935.

Colom, B, A Herms, M W J Hall, S C Dentro, C King, R K Sood, M P Alcolea, G Piedrafita, D Fernandez-Antoran, S H Ong, J C Fowler, K T Mahbubani, K Saeb-Parsy, M Gerstung, B A Hall and P H Jones (Oct. 2021). 'Mutant clones in normal epithelium outcompete and eliminate emerging tumours'. en. In: *Nature* 598.7881, pp. 510–514.

Conwill, Arolyn, Anne C Kuan, Ravalika Damerla, Alexandra J Poret, Jacob S Baker, A Delphine Tripp, Eric J Alm and Tami D Lieberman (Feb. 2022). 'Anatomy promotes neutral coexistence of strains in the human skin microbiome'. en. In: *Cell Host Microbe* 30.2, 171–182.e7.

Cords, Lena, Sandra Tietscher, Tobias Anzeneder, Claus Langwieder, Martin Rees, Natalie de Souza and Bernd Bodenmiller (July 2023). 'Cancer-associated fibroblast classification in single-cell and spatial proteomics data'. en. In: *Nat. Commun.* 14.1, p. 4294.

Coskun, Ahmet F and Long Cai (Aug. 2016). 'Dense transcript profiling in single cells by image correlation decoding'. en. In: *Nat. Methods* 13.8, pp. 657–660.

Costa, Ana, Yann Kieffer, Alix Scholer-Dahirel, Floriane Pelon, Brigitte Bourachot, Melissa Cardon, Philemon Sirven, Ilaria Magagna, Laetitia Fuhrmann, Charles Bernard, Claire Bonneau, Maria Kondratova, Inna Kuperstein, Andrei Zinovyev, Anne-Marie Givel, Maria-Carla Parrini, Vassili Soumelis, Anne Vincent-Salomon and Fatima Mechta-Grigoriou (Mar. 2018). 'Fibroblast Heterogeneity and Immunosuppressive Environment in Human Breast Cancer'. en. In: *Cancer Cell* 33.3, 463–479.e10.

Coudray, Nicolas, Paolo Santiago Ocampo, Theodore Sakellaropoulos, Navneet Narula, Matija Snuderl, David Fenyö, Andre L Moreira, Narges Razavian and Aristotelis Tsirigos (Oct. 2018). 'Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning'. en. In: *Nat. Med.* 24.10, pp. 1559–1567.

Cowell, Catherine F, Britta Weigelt, Rita A Sakr, Charlotte K Y Ng, James Hicks, Tari A King and Jorge S Reis-Filho (Oct. 2013). 'Progression from ductal carcinoma in situ to invasive breast cancer: revisited'. en. In: *Mol. Oncol.* 7.5, pp. 859–869.

Cronin, Kathleen A, Susan Scott, Albert U Firth, Hyuna Sung, S Jane Henley, Recinda L Sherman, Rebecca L Siegel, Robert N Anderson, Betsy A Kohler, Vicki B Benard, Serban Negoita, Charles Wiggins, William G Cance and Ahmedin Jemal (Dec. 2022). 'Annual report to the nation on the status of cancer, part 1: National cancer statistics'. en. In: *Cancer* 128.24, pp. 4251–4284.

Cui Zhou, Daniel, Reyka G Jayasinghe, Siqi Chen, John M Herndon, Michael D Iglesia, Pooja Navale, Michael C Wendl, Wagma Caravan, Kazuhito Sato, Erik Storrs, Chia-Kuei Mo, Jingxian Liu, Austin N Southard-Smith, Yige Wu, Nataly Naser Al Deen, John M Baer, Robert S Fulton, Matthew A Wyczalkowski, Ruiyang Liu, Catrina C Fronick, Lucinda A Fulton, Andrew Shinkle, Lisa Thammavong, Houxiang Zhu, Hua Sun, Liang-Bo Wang, Yize Li, Chong Zuo, Joshua F McMichael, Sherri R Davies, Elizabeth L Appelbaum, Keenan J Robbins, Sara E Chasnoff, Xiaolu Yang, Ashley N Reeb, Clara Oh, Mamatha Serasanambati, Preet Lal, Rajees Varghese, Jay R Mashl, Jennifer Ponce, Nadezhda V Terekhanova, Lijun Yao, Fang Wang, Lijun Chen, Michael Schnaubelt, Rita Jui-Hsien Lu, Julie K Schwarz, Sidharth V Puram, Albert H Kim, Sheng-Kwei Song, Kooresh I Shoghi, Ken S Lau, Tao Ju, Ken Chen, Deyali Chatterjee, William G Hawkins, Hui Zhang, Samuel Achilefu, Milan G Chheda, Stephen T Oh, William E Gillanders, Feng Chen, David G DeNardo, Ryan C Fields and Li Ding (Sept. 2022). 'Spatially restricted drivers and transitional cell populations cooperate with the microenvironment in untreated and chemo-resistant pancreatic cancer'. en. In: *Nat. Genet.* 54.9, pp. 1390–1405.

Curtis, Christina, Sohrab P Shah, Suet-Feung Chin, Gulisa Turashvili, Oscar M Rueda, Mark J Dunning, Doug Speed, Andy G Lynch, Shamith Samarajiwa, Yinyin Yuan, Stefan Gräf, Gavin Ha, Gholamreza Haffari, Ali Bashashati, Roslin Russell, Steven McKinney, METABRIC Group, Anita Langerød, Andrew Green, Elena Provenzano, Gordon Wishart, Sarah Pinder, Peter Watson, Florian Markowetz, Leigh Murphy, Ian Ellis, Arnie Purushotham, Anne-Lise Børresen-Dale, James D Brenton, Simon Tavaré, Carlos Caldas and Samuel Aparicio (Apr. 2012). 'The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups'. en. In: *Nature* 486.7403, pp. 346–352.

Dai, Min, Xiaobing Pei and Xiu-Jie Wang (Mar. 2022). 'Accurate and fast cell marker gene identification with COSG'. en. In: *Brief. Bioinform.* 23.2.

Danenberg, Esther, Helen Bardwell, Vito R T Zanotelli, Elena Provenzano, Suet-Feung Chin, Oscar M Rueda, Andrew Green, Emad Rakha, Samuel Aparicio, Ian O Ellis, Bernd Bodenmiller, Carlos Caldas and H Raza Ali (May 2022). 'Breast tumor microenvironment structures are associated with genomic features and clinical outcome'. en. In: *Nat. Genet.* 54.5, pp. 660–669.

Darmanis, Spyros, Steven A Sloan, Derek Croote, Marco Mignardi, Sophia Chernikova, Peyman Samghababi, Ye Zhang, Norma Neff, Mark Kowarsky, Christine Caneda, Gordon Li, Steven D Chang, Ian David Connolly, Yingmei Li, Ben A Barres, Melanie Hayden Gephart and Stephen R Quake (Oct. 2017). 'Single-Cell RNA-Seq Analysis of In-

filtrating Neoplastic Cells at the Migrating Front of Human Glioblastoma'. en. In: *Cell Rep.* 21.5, pp. 1399–1410.

Davidson, Sarah, Mirjana Efremova, Angela Riedel, Bidesh Mahata, Jhuma Pramanik, Jani Huuhtanen, Gozde Kar, Roser Vento-Tormo, Tzachi Hagai, Xi Chen, Muzlifah A Haniffa, Jacqueline D Shields and Sarah A Teichmann (May 2020). 'Single-Cell RNA Sequencing Reveals a Dynamic Stromal Niche That Supports Tumor Growth'. en. In: *Cell Rep.* 31.7, p. 107628.

De Mattos-Arruda, Leticia, Stephen-John Sammut, Edith M Ross, Rachael Bashford-Rogers, Erez Greenstein, Havell Markus, Sandro Morganella, Yvonne Teng, Yosef Maruvka, Bernard Pereira, Oscar M Rueda, Suet-Feung Chin, Tania Contente-Cuomo, Regina Mayor, Alexandra Arias, H Raza Ali, Wei Cope, Daniel Tiezzi, Aliakbar Dariush, Tauanne Dias Amarante, Dan Reshef, Nikaoly Ciriaco, Elena Martinez-Saez, Vicente Peg, Santiago Ramon Y Cajal, Javier Cortes, George Vassiliou, Gad Getz, Serena Nik-Zainal, Muhammed Murtaza, Nir Friedman, Florian Markowetz, Joan Seoane and Carlos Caldas (May 2019). 'The Genomic and Immune Landscapes of Lethal Metastatic Breast Cancer'. en. In: *Cell Rep.* 27.9, 2690–2708.e10.

DeNardo, David G and Brian Ruffell (June 2019). 'Macrophages as regulators of tumour immunity and immunotherapy'. en. In: *Nat. Rev. Immunol.* 19.6, pp. 369–382.

Deng, Yanxiang, Marek Bartosovic, Petra Kukanja, Di Zhang, Yang Liu, Graham Su, Archibald Enninful, Zhiliang Bai, Gonçalo Castelo-Branco and Rong Fan (Feb. 2022). 'Spatial-CUT&Tag: Spatially resolved chromatin modification profiling at the cellular level'. en. In: *Science* 375.6581, pp. 681–686.

Deng, Yanxiang, Marek Bartosovic, Sai Ma, Di Zhang, Petra Kukanja, Yang Xiao, Graham Su, Yang Liu, Xiaoyu Qin, Gorazd B Rosoklija, Andrew J Dwork, J John Mann, Mina L Xu, Stephanie Halene, Joseph E Craft, Kam W Leong, Maura Boldrini, Gonçalo Castelo-Branco and Rong Fan (Sept. 2022). 'Spatial profiling of chromatin accessibility in mouse and human tissues'. en. In: *Nature* 609.7926, pp. 375–383.

Dentro, Stefan C, Ignaty Leshchiner, Kerstin Haase, Maxime Tarabichi, Jeff Wintersinger, Amit G Deshwar, Kaixian Yu, Yulia Rubanova, Geoff Macintyre, Jonas Demeulemeester, Ignacio Vázquez-García, Kortine Kleinheinz, Dimitri G Livitz, Salem Malikic, Nilgun Donmez, Subhajit Sengupta, Pavana Anur, Clemency Jolly, Marek Cmero, Daniel Rosebrock, Steven E Schumacher, Yu Fan, Matthew Fittall, Ruben M Drews, Xiaotong Yao, Thomas B K Watkins, Juhee Lee, Matthias Schlesner, Hongtu Zhu, David J Adams, Nicholas McGranahan, Charles Swanton, Gad Getz, Paul C Boutros, Marcin Imielinski, Rameen Beroukhim, S Cenk Sahinalp, Yuan Ji, Martin Peifer, Inigo Martincorena, Florian Markowetz, Ville Mustonen, Ke Yuan, Moritz Gerstung, Paul T Spellman, Wenyi Wang, Quaid D Morris, David C Wedge, Peter Van Loo and PCAWG Evolution and Heterogeneity Working Group and the PCAWG Consortium (Apr. 2021). 'Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes'. en. In: *Cell* 184.8, 2239–2254.e39.

Dentro, Stefan C, David C Wedge and Peter Van Loo (Aug. 2017). 'Principles of Reconstructing the Subclonal Architecture of Cancers'. en. In: *Cold Spring Harb. Perspect. Med.* 7.8.

Dhainaut, Maxime, Samuel A Rose, Guray Akturk, Aleksandra Wroblewska, Sebastian R Nielsen, Eun Sook Park, Mark Buckup, Vladimir Roudko, Luisanna Pia, Robert Sweeney, Jessica Le Berichel, C Matthias Wilk, Anela Bektesevic, Brian H Lee, Nina Bhardwaj, Adeeb H Rahman, Alessia Baccarini, Sacha Gnjatic, Dana Pe'er, Miriam Merad and Brian D Brown (Mar. 2022). 'Spatial CRISPR genomics identifies regulators of the tumor microenvironment'. en. In: *Cell* 185.7, 1223–1239.e20.

Diamantopoulou, Zoi, Francesc Castro-Giner, Fabienne Dominique Schwab, Christiane Foerster, Massimo Saini, Selina Budinjas, Karin Strittmatter, Ilona Krol, Bettina Seifert, Viola Heinzelmann-Schwarz, Christian Kurzeder, Christoph Rochlitz, Marcus Vetter, Walter Paul Weber and Nicola Aceto (July 2022). 'The metastatic spread of breast cancer accelerates during sleep'. en. In: *Nature* 607.7917, pp. 156–162.

Dobzhansky, Theodosius (Mar. 1973). 'Nothing in biology makes sense except in the light of evolution'. en. In: *Am. Biol. Teach.* 35.3, pp. 125–129.

Dong, Kangning and Shihua Zhang (Apr. 2022). 'Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder'. en. In: *Nat. Commun.* 13.1, p. 1739.

Dries, Ruben, Jiaji Chen, Natalie Del Rossi, Mohammed Muzamil Khan, Adriana Sistig and Guo-Cheng Yuan (Oct. 2021). 'Advances in spatial transcriptomic data analysis'. en. In: *Genome Res.* 31.10, pp. 1706–1718.

Duane, Simon, A D Kennedy, Brian J Pendleton and Duncan Roweth (Sept. 1987). 'Hybrid Monte Carlo'. In: *Phys. Lett. B* 195.2, pp. 216–222.

Dumitrascu, Bianca, Soledad Villar, Dustin G Mixon and Barbara E Engelhardt (Feb. 2021). 'Optimal marker gene selection for cell type discrimination in single cell analyses'. en. In: *Nat. Commun.* 12.1, p. 1186.

Early Breast Cancer Trialists' Collaborative Group (EBCTCG) (2005). 'Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials'. en. In: *Lancet* 365.9472, pp. 1687–1717.

Ellis, Peter, Luiza Moore, Mathijs A Sanders, Timothy M Butler, Simon F Brunner, Henry Lee-Six, Robert Osborne, Ben Farr, Tim H H Coorens, Andrew R J Lawson, Alex Cagan, Mike R Stratton, Inigo Martincorena and Peter J Campbell (Feb. 2021). 'Reliable detection of somatic mutations in solid tissues by laser-capture microdissection and low-input DNA sequencing'. en. In: *Nat. Protoc.* 16.2, pp. 841–871.

Elston, C W and I O Ellis (Nov. 1991). 'Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up'. en. In: *Histopathology* 19.5, pp. 403–410.

Elyanow, Rebecca, Ron Zeira, Max Land and Benjamin J Raphael (Mar. 2021). 'STARCH: copy number and clone inference from spatial transcriptomics data'. en. In: *Phys. Biol.* 18.3, p. 035001.

Emmert-Buck, M R, R F Bonner, P D Smith, R F Chuaqui, Z Zhuang, S R Goldstein, R A Weiss and L A Liotta (Nov. 1996). 'Laser capture microdissection'. en. In: *Science* 274.5289, pp. 998–1001.

Eng, Chee-Huat Linus, Michael Lawson, Qian Zhu, Ruben Dries, Noushin Koulena, Yodai Takei, Jina Yun, Christopher Cronin, Christoph Karp, Guo-Cheng Yuan and Long Cai (Apr. 2019). 'Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH'. en. In: *Nature* 568.7751, pp. 235–239.

Epstein, Jonathan I, Lars Egevad, Mahul B Amin, Brett Delahunt, John R Srigley, Peter A Humphrey and Grading Committee (Feb. 2016). 'The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System'. en. In: *Am. J. Surg. Pathol.* 40.2, pp. 244–252.

Erickson, Andrew, Mengxiao He, Emelie Berglund, Maja Marklund, Reza Mirzazadeh, Niklas Schultz, Linda Kvastad, Alma Andersson, Ludvig Bergenstråhle, Joseph Bergenstråhle, Ludvig Larsson, Leire Alonso Galicia, Alia Shamikh, Elisa Basmaci, Teresita Díaz De Ståhl, Timothy Rajakumar, Dimitrios Doultsinos, Kim Thrane, Andrew L Ji, Paul A Khavari, Firaz Tarish, Anna Tanoglidi, Jonas Maaskola, Richard Colling, Tuomas Mirtti, Freddie C Hamdy, Dan J Woodcock, Thomas Helleday, Ian G Mills, Alastair D Lamb and Joakim Lundeberg (Aug. 2022). 'Spatially resolved clonal copy number alterations in benign and malignant tissue'. en. In: *Nature* 608.7922, pp. 360–367.

Fane, Mitchell E, Yash Chhabra, Gretchen M Alicea, Devon A Maranto, Stephen M Douglass, Marie R Webster, Vito W Rebecca, Gloria E Marino, Filipe Almeida, Brett L Ecker, Daniel J Zabransky, Laura Hüser, Thomas Beer, Hsin-Yao Tang, Andrew Kossenkov, Meenhard Herlyn, David W Speicher, Wei Xu, Xiaowei Xu, Elizabeth M Jaffee, Julio A Aguirre-Ghiso and Ashani T Weeraratna (June 2022). 'Stromal changes in the aged lung induce an emergence from melanoma dormancy'. en. In: *Nature* 606.7913, pp. 396–405.

Färkkilä, Anniina, Doga C Gulhan, Julia Casado, Connor A Jacobson, Huy Nguyen, Bose Kochupurakkal, Zoltan Maliga, Clarence Yapp, Yu-An Chen, Denis Schapiro, Yinghui Zhou, Julie R Graham, Bruce J Dezube, Pamela Munster, Sandro Santagata, Elizabeth Garcia, Scott Rodig, Ana Lako, Dipanjan Chowdhury, Geoffrey I Shapiro, Ursula A Matulonis, Peter J Park, Sampsa Hautaniemi, Peter K Sorger, Elizabeth M Swisher, Alan D D'Andrea and Panagiotis A Konstantinopoulos (Mar. 2020). 'Immunogenomic profiling determines responses to combined PARP and PD-1 inhibition in ovarian cancer'. en. In: *Nat. Commun.* 11.1, p. 1459.

Fearon, E R and B Vogelstein (June 1990). 'A genetic model for colorectal tumorigenesis'. en. In: *Cell* 61.5, pp. 759–767.

Ferrari, Anthony, Anne Vincent-Salomon, Xavier Pivot, Anne-Sophie Sertier, Emilie Thomas, Laurie Tonon, Sandrine Boyault, Eskeatnaf Mulugeta, Isabelle Treilleux, Gaëtan MacGrogan, Laurent Arnould, Janice Kielbassa, Vincent Le Texier, Hélène Blanché, Jean-François Deleuze, Jocelyne Jacquemier, Marie-Christine Mathieu,

Frédérique Penault-Llorca, Frédéric Bibeau, Odette Mariani, Cécile Mannina, Jean-Yves Pierga, Olivier Trédan, Thomas Bachelot, Hervé Bonnefoi, Gilles Romieu, Pierre Fumoleau, Suzette Delaloge, Maria Rios, Jean-Marc Ferrero, Carole Tarpin, Catherine Bouteille, Fabien Calvo, Ivo Glynne Gut, Marta Gut, Sancha Martin, Serena Nik-Zainal, Michael R Stratton, Iris Pauporté, Pierre Saintigny, Daniel Birnbaum, Alain Viari and Gilles Thomas (July 2016). 'A whole-genome sequence and transcriptome perspective on HER2-positive breast cancers'. en. In: *Nat. Commun.* 7, p. 12222.

Fidler, Isaiah J (June 2003). 'The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited'. en. In: *Nat. Rev. Cancer* 3.6, pp. 453–458.

Fischer, David S, Anna C Schaar and Fabian J Theis (Mar. 2023). 'Modeling intercellular communication in tissues using spatial graphs of cells'. en. In: *Nat. Biotechnol.* 41.3, pp. 332–336.

Fisher, B, E Montague, C Redmond, B Barton, D Borland, E R Fisher, M Deutsch, G Schwarz, R Margolese, W Donegan, H Volk, C Konvolinka, B Gardner, I Cohn Jr, G Lesnick, A B Cruz, W Lawrence, T Nealon, H Butcher and R Lawton (June 1977). 'Comparison of radical mastectomy with alternative treatments for primary breast cancer. A first report of results from a prospective randomized clinical trial'. en. In: *Cancer* 39.6 Suppl, pp. 2827–2839.

Fowler, Joanna C, Charlotte King, Christopher Bryant, Michael W J Hall, Roshan Sood, Swee Hoe Ong, Eleanor Earp, David Fernandez-Antoran, Jonas Koeppel, Stefan C Dentro, David Shorthouse, Amer Durrani, Kate Fife, Edward Rytina, Doreen Milne, Amit Roshan, Krishnaa Mahububani, Kourosh Saeb-Parsy, Benjamin A Hall, Moritz Gerstung and Philip H Jones (Feb. 2021). 'Selection of Oncogenic Mutant Clones in Normal Human Skin Varies with Body Site'. en. In: *Cancer Discov.* 11.2, pp. 340–361.

Frean, Marcus, Paul B Rainey and Arne Traulsen (July 2013). 'The effect of population structure on the rate of evolution'. en. In: *Proc. Biol. Sci.* 280.1762, p. 20130211.

Fu, Yu, Alexander W Jung, Ramon Viñas Torne, Santiago Gonzalez, Harald Vöhringer, Artem Shmatko, Lucy R Yates, Mercedes Jimenez-Linan, Luiza Moore and Moritz Gerstung (Aug. 2020). 'Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis'. en. In: *Nat Cancer* 1.8, pp. 800–810.

Gaglia, Giorgio, Sheheryar Kabraji, Danae Rammos, Yang Dai, Ana Verma, Shu Wang, Caitlin E Mills, Mirra Chung, Johann S Bergholz, Shannon Coy, Jia-Ren Lin, Rinath Jeselsohn, Otto Metzger, Eric P Winer, Deborah A Dillon, Jean J Zhao, Peter K Sorger and Sandro Santagata (Mar. 2022). 'Temporal and spatial topography of cell proliferation in cancer'. en. In: *Nat. Cell Biol.* 24.3, pp. 316–326.

Gallaher, Jill A, Joel S Brown and Alexander R A Anderson (Feb. 2019). 'The impact of proliferation-migration tradeoffs on phenotypic evolution in cancer'. en. In: *Sci. Rep.* 9.1, p. 2425.

Galon, Jérôme, Anne Costes, Fatima Sanchez-Cabo, Amos Kirilovsky, Bernhard Mlecnik, Christine Lagorce-Pagès, Marie Tosolini, Matthieu Camus, Anne Berger, Philippe Wind, Franck Zinzindohoué, Patrick Bruneval, Paul-Henri Cugnenc, Zlatko Trajanoski, Wolf-Herman Fridman and Franck Pagès (Sept. 2006). 'Type, density, and loc-

ation of immune cells within human colorectal tumors predict clinical outcome'. en. In: *Science* 313.5795, pp. 1960–1964.

Galon, Jérôme, Bernhard Mlecnik, Gabriela Bindea, Helen K Angell, Anne Berger, Christine Lagorce, Alessandro Lugli, Inti Zlobec, Arndt Hartmann, Carlo Bifulco, Iris D Nagtegaal, Richard Palmqvist, Giuseppe V Masucci, Gerardo Botti, Fabiana Tatangelo, Paolo Delrio, Michele Maio, Luigi Laghi, Fabio Grizzi, Martin Asslaber, Corrado D'Arrigo, Fernando Vidal-Vanaclocha, Eva Zavadova, Lotfi Chouchane, Pamela S Ohashi, Sara Hafezi-Bakhtiari, Bradly G Wouters, Michael Roehrl, Linh Nguyen, Yutaka Kawakami, Shoichi Hazama, Kiyotaka Okuno, Shuji Ogino, Peter Gibbs, Paul Waring, Noriyuki Sato, Toshihiko Torigoe, Kyogo Itoh, Prabhu S Patel, Shilin N Shukla, Yili Wang, Scott Kopetz, Frank A Sinicrope, Viorel Scripcariu, Paolo A Ascierto, Francesco M Marincola, Bernard A Fox and Franck Pagès (Jan. 2014). 'Towards the introduction of the 'Immunoscore' in the classification of malignant tumours'. en. In: *J. Pathol.* 232.2, pp. 199–209.

Gao, Ruli, Alexander Davis, Thomas O McDonald, Emi Sei, Xiuqing Shi, Yong Wang, Pei-Ching Tsai, Anna Casasent, Jill Waters, Hong Zhang, Funda Meric-Bernstam, Franziska Michor and Nicholas E Navin (Oct. 2016). 'Punctuated copy number evolution and clonal stasis in triple-negative breast cancer'. en. In: *Nat. Genet.* 48.10, pp. 1119–1130.

Gataric, Milana, Jun Sung Park, Tong Li, Vasyl Vaskivskyi, Jessica Svedlund, Carina Strell, Kenny Roberts, Mats Nilsson, Lucy R Yates, Omer Bayraktar and Moritz Gerstung (Oct. 2021). 'PoSTcode: Probabilistic image-based spatial transcriptomics decoder'. en.

Gatenbee, Chandler D, Ann-Marie Baker, Ryan O Schenck, Maximilian Strobl, Jeffrey West, Margarida P Neves, Sara Yakub Hasan, Eszter Lakatos, Pierre Martinez, William C H Cross, Marnix Jansen, Manuel Rodriguez-Justo, Christopher J Whelan, Andrea Sottoriva, Simon Leedham, Mark Robertson-Tessi, Trevor A Graham and Alexander R A Anderson (Apr. 2022). 'Immunosuppressive niche engineering at the onset of human colorectal cancer'. en. In: *Nat. Commun.* 13.1, p. 1798.

Geman, Stuart and Donald Geman (Nov. 1984). 'Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images'. In: *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-6.6, pp. 721–741.

Genovese, Giulio, Anna K Kähler, Robert E Handsaker, Johan Lindberg, Samuel A Rose, Samuel F Bakhoum, Kimberly Chambert, Eran Mick, Benjamin M Neale, Menachem Fromer, Shaun M Purcell, Oscar Svantesson, Mikael Landén, Martin Höglund, Sören Lehmann, Stacey B Gabriel, Jennifer L Moran, Eric S Lander, Patrick F Sullivan, Pamela Sklar, Henrik Grönberg, Christina M Hultman and Steven A McCarroll (Dec. 2014). 'Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence'. en. In: *N. Engl. J. Med.* 371.26, pp. 2477–2487.

Gerdes, Michael J, Christopher J Sevinsky, Anup Sood, Sudeshna Adak, Musodiq O Bello, Alexander Bordwell, Ali Can, Alex Corwin, Sean Dinn, Robert J Filkins, Denise Hollman, Vidya Kamath, Sireesha Kaanumalle, Kevin Kenny, Melinda Larsen, Michael Lazare, Qing Li, Christina Lowes, Colin C McCulloch, Elizabeth McDonough, Michael C Montalto, Zhengyu Pang, Jens Rittscher, Alberto Santamaria-Pang, Brion D Sarachan,

Maximilian L Seel, Antti Seppo, Kashan Shaikh, Yunxia Sui, Jingyu Zhang and Fiona Ginty (July 2013). 'Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue'. en. In: *Proc. Natl. Acad. Sci. U. S. A.* 110.29, pp. 11982–11987.

Gerlinger, Marco, Stuart Horswell, James Larkin, Andrew J Rowan, Max P Salm, Ignacio Varela, Rosalie Fisher, Nicholas McGranahan, Nicholas Matthews, Claudio R Santos, Pierre Martinez, Benjamin Phillimore, Sharmin Begum, Adam Rabinowitz, Bradley Spencer-Dene, Sakshi Gulati, Paul A Bates, Gordon Stamp, Lisa Pickering, Martin Gore, David L Nicol, Steven Hazell, P Andrew Futreal, Aengus Stewart and Charles Swanton (Mar. 2014). 'Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing'. en. In: *Nat. Genet.* 46.3, pp. 225–233.

Gerlinger, Marco, Andrew J Rowan, Stuart Horswell, M Math, James Larkin, David Endesfelder, Eva Gronroos, Pierre Martinez, Nicholas Matthews, Aengus Stewart, Patrick Tarpey, Ignacio Varela, Benjamin Phillimore, Sharmin Begum, Neil Q McDonald, Adam Butler, David Jones, Keiran Raine, Calli Latimer, Claudio R Santos, Mahrokh Nohadani, Aron C Eklund, Bradley Spencer-Dene, Graham Clark, Lisa Pickering, Gordon Stamp, Martin Gore, Zoltan Szallasi, Julian Downward, P Andrew Futreal and Charles Swanton (Mar. 2012). 'Intratumor heterogeneity and branched evolution revealed by multiregion sequencing'. en. In: *N. Engl. J. Med.* 366.10, pp. 883–892.

Gerstung, Moritz, Clemency Jolly, Ignaty Leshchiner, Stefan C Dentro, Santiago Gonzalez, Daniel Rosebrock, Thomas J Mitchell, Yulia Rubanova, Pavana Anur, Kaixian Yu, Maxime Tarabichi, Amit Deshwar, Jeff Wintersinger, Kortine Kleinheinz, Ignacio Vázquez-García, Kerstin Haase, Lara Jerman, Subhajit Sengupta, Geoff Macintyre, Salem Malikic, Nilgun Donmez, Dimitri G Livitz, Marek Cmero, Jonas Demeulemeester, Steven Schumacher, Yu Fan, Xiaotong Yao, Juhee Lee, Matthias Schlesner, Paul C Boutros, David D Bowtell, Hongtu Zhu, Gad Getz, Marcin Imielinski, Rameen Beroukhim, S Cenk Sahinalp, Yuan Ji, Martin Peifer, Florian Markowetz, Ville Mustonen, Ke Yuan, Wenyi Wang, Quaid D Morris, PCAWG Evolution & Heterogeneity Working Group, Paul T Spellman, David C Wedge, Peter Van Loo and PCAWG Consortium (Feb. 2020). 'The evolutionary history of 2,658 cancers'. en. In: *Nature* 578.7793, pp. 122–128.

Ghajar, Cyrus M, Héctor Peinado, Hidetoshi Mori, Irina R Matei, Kimberley J Evason, Hélène Brazier, Dena Almeida, Antonius Koller, Katherine A Hajjar, Didier Y R Stainier, Emily I Chen, David Lyden and Mina J Bissell (July 2013). 'The perivascular niche regulates breast tumour dormancy'. en. In: *Nat. Cell Biol.* 15.7, pp. 807–817.

Giesen, Charlotte, Hao A O Wang, Denis Schapiro, Nevena Zivanovic, Andrea Jacobs, Bodo Hattendorf, Peter J Schüffler, Daniel Grolimund, Joachim M Buhmann, Simone Brandt, Zsuzsanna Varga, Peter J Wild, Detlef Günther and Bernd Bodenmiller (Apr. 2014). 'Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry'. en. In: *Nat. Methods* 11.4, pp. 417–422.

Going, James J and David F Moffat (May 2004). 'Escaping from Flatland: clinical and biological aspects of human mammary duct anatomy in three dimensions'. en. In: *J. Pathol.* 203.1, pp. 538–544.

Goltsev, Yury, Nikolay Samusik, Julia Kennedy-Darling, Salil Bhate, Matthew Hale, Gustavo Vazquez, Sarah Black and Garry P Nolan (Aug. 2018). 'Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging'. en. In: *Cell* 174.4, 968–981.e15.

Gonzalez-Perez, Abel, Christian Perez-Llamas, Jordi Deu-Pons, David Tamborero, Michael P Schroeder, Alba Jene-Sanz, Alberto Santos and Nuria Lopez-Bigas (Nov. 2013). 'IntOGen-mutations identifies cancer drivers across tumor types'. en. In: *Nat. Methods* 10.11, pp. 1081–1082.

González-Silva, Laura, Laura Quevedo and Ignacio Varela (Jan. 2020). 'Tumor Functional Heterogeneity Unraveled by scRNA-seq Technologies'. en. In: *Trends Cancer Res.* 6.1, pp. 13–19.

Goodman, Aaron M, David Piccioni, Shumei Kato, Amélie Boichard, Huan-You Wang, Garrett Frampton, Scott M Lippman, Caitlin Connelly, David Fabrizio, Vincent Miller, Jason K Sicklick and Razelle Kurzrock (Sept. 2018). 'Prevalence of PDL1 Amplification and Preliminary Response to Immune Checkpoint Blockade in Solid Tumors'. en. In: *JAMA Oncol* 4.9, pp. 1237–1244.

Goswami, Sangeeta, Swetha Anandhan, Deblina Raychaudhuri and Padmanee Sharma (Feb. 2023). 'Myeloid cell-targeted therapies for solid tumours'. en. In: *Nat. Rev. Immunol.* 23.2, pp. 106–120.

Gouin 3rd, Kenneth H, Nathan Ing, Jasmine T Plummer, Charles J Rosser, Bassem Ben Cheikh, Catherine Oh, Stephanie S Chen, Keith Syson Chan, Hideki Furuya, Warren G Tourtellotte, Simon R V Knott and Dan Theodorescu (Aug. 2021). 'An N-Cadherin 2 expressing epithelial cell subpopulation predicts response to surgery, chemotherapy and immunotherapy in bladder cancer'. en. In: *Nat. Commun.* 12.1, p. 4906.

Grabski, Isabella N and Rafael A Irizarry (Oct. 2022). 'A probabilistic gene expression barcode for annotation of cell types from single-cell RNA-seq data'. en. In: *Biostatistics* 23.4, pp. 1150–1164.

Graeber, T G, C Osmanian, T Jacks, D E Housman, C J Koch, S W Lowe and A J Giaccia (Jan. 1996). 'Hypoxia-mediated selection of cells with diminished apoptotic potential in solid tumours'. en. In: *Nature* 379.6560, pp. 88–91.

Greenman, Chris, Richard Wooster, P Andrew Futreal, Michael R Stratton and Douglas F Easton (Aug. 2006). 'Statistical analysis of pathogenicity of somatic mutations in cancer'. en. In: *Genetics* 173.4, pp. 2187–2198.

Greenough, Robert B (Dec. 1925). 'Varying degrees of malignancy in cancer of the breast'. en. In: *J. Cancer Res.* 9.4, pp. 453–463.

Grossmann, Sebastian, Yvette Hooks, Laura Wilson, Luiza Moore, Laura O'Neill, Iñigo Martincorena, Thierry Voet, Michael R Stratton, Rakesh Heer and Peter J Campbell (July 2021). 'Development, maturation, and maintenance of human prostate inferred from somatic mutations'. en. In: *Cell Stem Cell* 28.7, 1262–1274.e5.

Grundberg, Ida, Sara Kiflemariam, Marco Mignardi, Juliana Imgenberg-Kreuz, Karolina Edlund, Patrick Micke, Magnus Sundström, Tobias Sjöblom, Johan Botling and Mats

Nilsson (Dec. 2013). 'In situ mutation detection and visualization of intratumor heterogeneity for cancer research and diagnostics'. en. In: *Oncotarget* 4.12, pp. 2407–2418.

Grünwald, Barbara T, Antoine Devisme, Geoffroy Andrieux, Foram Vyas, Kazeera Aliar, Curtis W McCloskey, Andrew Macklin, Gun Ho Jang, Robert Denroche, Joan Miguel Romero, Prashant Bavi, Peter Bronsert, Faiyaz Notta, Grainne O'Kane, Julie Wilson, Jennifer Knox, Laura Tamblyn, Molly Udaskin, Nikolina Radulovich, Sandra E Fischer, Melanie Boerries, Steven Gallinger, Thomas Kislinger and Rama Khokha (Oct. 2021). 'Spatially confined sub-tumor microenvironments in pancreatic cancer'. en. In: *Cell* 184.22, 5577–5592.e18.

Gundem, Gunes, Peter Van Loo, Barbara Kremeyer, Ludmil B Alexandrov, Jose M C Tubio, Elli Papaemmanuil, Daniel S Brewer, Heini M L Kallio, Gunilla Högnäs, Matti Annala, Kati Kivinummi, Victoria Goody, Calli Latimer, Sarah O'Meara, Kevin J Dawson, William Isaacs, Michael R Emmert-Buck, Matti Nykter, Christopher Foster, Zsofia Kote-Jarai, Douglas Easton, Hayley C Whitaker, ICGC Prostate Group, David E Neal, Colin S Cooper, Rosalind A Eeles, Tapio Visakorpi, Peter J Campbell, Ultan McDermott, David C Wedge and G Steven Bova (Apr. 2015). 'The evolutionary history of lethal metastatic prostate cancer'. en. In: *Nature* 520.7547, pp. 353–357.

Gupta, S K, A G Douglas-Jones, N Fenn, J M Morgan and R E Mansel (Nov. 1997). 'The clinical behavior of breast carcinoma is probably determined at the preinvasive stage (ductal carcinoma in situ)'. en. In: *Cancer* 80.9, pp. 1740–1745.

Gyllborg, Daniel, Christoffer Mattsson Langseth, Xiaoyan Qian, Eunkyoung Choi, Sergio Marco Salas, Markus M Hilscher, Ed S Lein and Mats Nilsson (Nov. 2020). 'Hybridization-based in situ sequencing (HybISS) for spatially resolved transcriptomics in human and mouse brain tissue'. en. In: *Nucleic Acids Res.* 48.19, e112.

Hall, J M, M K Lee, B Newman, J E Morrow, L A Anderson, B Huey and M C King (Dec. 1990). 'Linkage of early-onset familial breast cancer to chromosome 17q21'. en. In: *Science* 250.4988, pp. 1684–1689.

Hammerl, Dora, John W M Martens, Mieke Timmermans, Marcel Smid, Anita M Trapman-Jansen, Renée Foekens, Olga I Isaeva, Leonie Voorwerk, Hayri E Balcioglu, Rebecca Wijers, Iris Nederlof, Roberto Salgado, Hugo Horlings, Marleen Kok and Reno Debets (Sept. 2021). 'Spatial immunophenotypes predict response to anti-PD1 treatment and capture distinct paths of T cell evasion in triple negative breast cancer'. en. In: *Nat. Commun.* 12.1, p. 5668.

Hanahan, D and R A Weinberg (Jan. 2000). 'The hallmarks of cancer'. en. In: *Cell* 100.1, pp. 57–70.

Hanahan, Douglas (Jan. 2022). 'Hallmarks of Cancer: New Dimensions'. en. In: *Cancer Discov.* 12.1, pp. 31–46.

Hanahan, Douglas and Robert A Weinberg (Mar. 2011). 'Hallmarks of cancer: the next generation'. en. In: *Cell* 144.5, pp. 646–674.

Handsaker, Robert E, Vanessa Van Doren, Jennifer R Berman, Giulio Genovese, Seva Kashin, Linda M Boettger and Steven A McCarroll (Mar. 2015). 'Large multiallelic copy number variations in humans'. en. In: *Nat. Genet.* 47.3, pp. 296–303.

Hao, Yuhan, Stephanie Hao, Erica Andersen-Nissen, William M Mauck 3rd, Shiwei Zheng, Andrew Butler, Maddie J Lee, Aaron J Wilk, Charlotte Darby, Michael Zager, Paul Hoffman, Marlon Stoeckius, Efthymia Papalexi, Eleni P Mimitou, Jaison Jain, Avi Srivastava, Tim Stuart, Lamar M Fleming, Bertrand Yeung, Angela J Rogers, Juliana M McElrath, Catherine A Blish, Raphael Gottardo, Peter Smibert and Rahul Satija (June 2021). 'Integrated analysis of multimodal single-cell data'. en. In: *Cell* 184.13, 3573–3587.e29.

Harris, Adrian L (Jan. 2002). 'Hypoxia–a key regulatory factor in tumour growth'. en. In: *Nat. Rev. Cancer* 2.1, pp. 38–47.

Hastings, W K (1970). 'Monte Carlo Sampling Methods Using Markov Chains and Their Applications'. In: *Biometrika* 57.1, pp. 97–109.

He, S (2022). 'High-plex multiomic analysis in FFPE at subcellular level by spatial molecular imaging'. In: *bioRxiv*.

Hegde, Priti S and Daniel S Chen (Jan. 2020). 'Top 10 Challenges in Cancer Immunotherapy'. en. In: *Immunity* 52.1, pp. 17–35.

Heide, Timon, Jacob Househam, George D Cresswell, Inmaculada Spiteri, Claire Lynn, Maximilian Mossner, Chris Kimberley, Javier Fernandez-Mateos, Bingjie Chen, Luis Zapata, Chela James, Iros Barozzi, Ketevan Chkhaidze, Daniel Nichol, Vinaya Gunasri, Alison Berner, Melissa Schmidt, Eszter Lakatos, Ann-Marie Baker, Helena Costa, Miriam Mitchinson, Rocco Piazza, Marnix Jansen, Giulio Caravagna, Daniele Ramazzotti, Darryl Shibata, John Bridgewater, Manuel Rodriguez-Justo, Luca Magnani, Trevor A Graham and Andrea Sottoriva (Nov. 2022). 'The co-evolution of the genome and epigenome in colorectal cancer'. en. In: *Nature* 611.7937, pp. 733–743.

Heide, Timon, Angela Maurer, Monika Eipel, Katrin Knoll, Mirja Geelvink, Juergen Veeck, Ruth Knuechel, Julius van Essen, Robert Stoehr, Arndt Hartmann, Janine Altmueller, Trevor A Graham and Nadine T Gaisa (June 2019). 'Multiregion human bladder cancer sequencing reveals tumour evolution, bladder cancer phenotypes and implications for targeted therapy'. en. In: *J. Pathol.* 248.2, pp. 230–242.

Helmink, Beth A, Sangeetha M Reddy, Jianjun Gao, Shaojun Zhang, Rafet Basar, Rohit Thakur, Keren Yizhak, Moshe Sade-Feldman, Jorge Blando, Guangchun Han, Vancheswaran Gopalakrishnan, Yuanxin Xi, Hao Zhao, Rodabe N Amaria, Hussein A Tawbi, Alex P Cogdill, Wenbin Liu, Valerie S LeBleu, Fernanda G Kugeratski, Sapna Patel, Michael A Davies, Patrick Hwu, Jeffrey E Lee, Jeffrey E Gershenwald, Anthony Lucci, Reetakshi Arora, Scott Woodman, Emily Z Keung, Pierre-Olivier Gaudreau, Alexandre Reuben, Christine N Spencer, Elizabeth M Burton, Lauren E Haydu, Alexander J Lazar, Roberta Zapassodi, Courtney W Hudgens, Deborah A Ledesma, Sufey Ong, Michael Bailey, Sarah Warren, Disha Rao, Oscar Krijgsman, Elisa A Rozeman, Daniel Peeper, Christian U Blank, Ton N Schumacher, Lisa H Butterfield, Monika A Zelazowska, Kevin M McBride, Raghu Kalluri, James Allison, Florent Petitprez, Wolf Herman Fridman, Catherine Sautès-Fridman, Nir Hacohen, Katayoun Rezvani, Padmanee Sharma, Michael T Tetzlaff, Linghua Wang and Jennifer A Wargo (Jan.

2020). 'B cells and tertiary lymphoid structures promote immunotherapy response'. en. In: *Nature* 577.7791, pp. 549–555.

Hickey, John W, Elizabeth K Neumann, Andrea J Radtke, Jeannie M Camarillo, Rebecca T Beuschel, Alexandre Albanese, Elizabeth McDonough, Julia Hatler, Anne E Wiblin, Jeremy Fisher, Josh Croteau, Eliza C Small, Anup Sood, Richard M Caprioli, R Michael Angelo, Garry P Nolan, Kwanghun Chung, Stephen M Hewitt, Ronald N Germain, Jeffrey M Spraggins, Emma Lundberg, Michael P Snyder, Neil L Kelleher and Sinem K Saka (Mar. 2022). 'Spatial mapping of protein composition and tissue organization: a primer for multiplexed antibody-based imaging'. en. In: *Nat. Methods* 19.3, pp. 284–295.

Hörl, David, Fabio Rojas Rusak, Friedrich Preusser, Paul Tillberg, Nadine Randel, Raghav K Chhetri, Albert Cardona, Philipp J Keller, Hartmann Harz, Heinrich Leonhardt, Mathias Treier and Stephan Preibisch (Sept. 2019). 'BigStitcher: reconstructing high-resolution image datasets of cleared and expanded samples'. en. In: *Nat. Methods* 16.9, pp. 870–874.

Househam, Jacob, Timon Heide, George D Cresswell, Inmaculada Spiteri, Chris Kimberley, Luis Zapata, Claire Lynn, Chela James, Maximilian Mossner, Javier Fernandez-Mateos, Alessandro Vinceti, Ann-Marie Baker, Calum Gabbutt, Alison Berner, Melissa Schmidt, Bingjie Chen, Eszter Lakatos, Vinaya Gunasri, Daniel Nichol, Helena Costa, Miriam Mitchinson, Daniele Ramazzotti, Benjamin Werner, Francesco Iorio, Marnix Jansen, Giulio Caravagna, Chris P Barnes, Darryl Shibata, John Bridgewater, Manuel Rodriguez-Justo, Luca Magnani, Andrea Sottoriva and Trevor A Graham (Nov. 2022). 'Phenotypic plasticity and genetic control in colorectal cancer evolution'. en. In: *Nature* 611.7937, pp. 744–753.

Hu, Zheng, Zan Li, Zhicheng Ma and Christina Curtis (July 2020). 'Multi-cancer analysis of clonality and the timing of systemic spread in paired primary tumors and metastases'. en. In: *Nat. Genet.* 52.7, pp. 701–708.

Humphries, Adam and Nicholas A Wright (June 2008). 'Colonic crypt organization and tumorigenesis'. en. In: *Nat. Rev. Cancer* 8.6, pp. 415–424.

ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium (Feb. 2020). 'Pan-cancer analysis of whole genomes'. en. In: *Nature* 578.7793, pp. 82–93.

International Cancer Genome Consortium et al. (Apr. 2010). 'International network of cancer genome projects'. en. In: *Nature* 464.7291, pp. 993–998.

Izar, Benjamin, Itay Tirosh, Elizabeth H Stover, Isaac Wakiro, Michael S Cuoco, Idan Alter, Christopher Rodman, Rachel Leeson, Mei-Ju Su, Parin Shah, Marcin Iwanicki, Sarah R Walker, Abhay Kanodia, Johannes C Melms, Shaolin Mei, Jia-Ren Lin, Caroline B M Porter, Michal Slyper, Julia Waldman, Livnat Jerby-Arnon, Orr Ashenberg, Titus J Brinker, Caitlin Mills, Meri Rogava, Sébastien Vigneau, Peter K Sorger, Levi A Garraway, Panagiotis A Konstantinopoulos, Joyce F Liu, Ursula Matulonis, Bruce E Johnson, Orit Rozenblatt-Rosen, Asaf Rotem and Aviv Regev (Aug. 2020). 'A single-cell landscape of high-grade serous ovarian cancer'. en. In: *Nat. Med.* 26.8, pp. 1271–1279.

Jackson, Hartland W, Jana R Fischer, Vito R T Zanotelli, H Raza Ali, Robert Mechera, Savas D Soysal, Holger Moch, Simone Muenst, Zsuzsanna Varga, Walter P Weber and Bernd Bodenmiller (Feb. 2020). 'The single-cell pathology landscape of breast cancer'. en. In: *Nature* 578.7796, pp. 615–620.

Jaiswal, Siddhartha, Pierre Fontanillas, Jason Flannick, Alisa Manning, Peter V Grauman, Brenton G Mar, R Coleman Lindsley, Craig H Mermel, Noel Burtt, Alejandro Chavez, John M Higgins, Vladislav Moltchanov, Frank C Kuo, Michael J Kluk, Brian Henderson, Leena Kinnunen, Heikki A Koistinen, Claes Ladenvall, Gad Getz, Adolfo Correa, Benjamin F Banahan, Stacey Gabriel, Sekar Kathiresan, Heather M Stringham, Mark I McCarthy, Michael Boehnke, Jaakko Tuomilehto, Christopher Haiman, Leif Groop, Gil Atzmon, James G Wilson, Donna Neuberg, David Altshuler and Benjamin L Ebert (Dec. 2014). 'Age-related clonal hematopoiesis associated with adverse outcomes'. en. In: *N. Engl. J. Med.* 371.26, pp. 2488–2498.

Jamal-Hanjani, Mariam, Gareth A Wilson, Nicholas McGranahan, Nicolai J Birkbak, Thomas B K Watkins, Selvaraju Veeriah, Seema Shafi, Diana H Johnson, Richard Mitter, Rachel Rosenthal, Max Salm, Stuart Horswell, Mickael Escudero, Nik Matthews, Andrew Rowan, Tim Chambers, David A Moore, Samra Turajlic, Hang Xu, Siow-Ming Lee, Martin D Forster, Tanya Ahmad, Crispin T Hiley, Christopher Abbosh, Mary Falzon, Elaine Borg, Teresa Marafioti, David Lawrence, Martin Hayward, Shyam Kolvekar, Nikolaos Panagiotopoulos, Sam M Janes, Ricky Thakrar, Asia Ahmed, Fiona Blackhall, Yvonne Summers, Rajesh Shah, Leena Joseph, Anne M Quinn, Phil A Crosbie, Babu Naidu, Gary Middleton, Gerald Langman, Simon Trotter, Marianne Nicolson, Hardy Remmen, Keith Kerr, Mahendran Chetty, Lesley Gomersall, Dean A Fennell, Apostolos Nakas, Sridhar Rathinam, Girija Anand, Sajid Khan, Peter Russell, Veni Ezhil, Babikir Ismail, Melanie Irvin-Sellers, Vineet Prakash, Jason F Lester, Malgorzata Kornaszewska, Richard Attanoos, Haydn Adams, Helen Davies, Stefan Dentro, Philippe Taniere, Brendan O'Sullivan, Helen L Lowe, John A Hartley, Natasha Iles, Harriet Bell, Yenting Ngai, Jacqui A Shaw, Javier Herrero, Zoltan Szallasi, Roland F Schwarz, Aengus Stewart, Sergio A Quezada, John Le Quesne, Peter Van Loo, Caroline Dive, Allan Hackshaw, Charles Swanton and TRACERx Consortium (June 2017). 'Tracking the Evolution of Non-Small-Cell Lung Cancer'. en. In: *N. Engl. J. Med.* 376.22, pp. 2109–2121.

Janiszewska, Michalina, Lin Liu, Vanessa Almendro, Yanan Kuang, Cloud Paweletz, Rita A Sakr, Britta Weigelt, Ariella B Hanker, Sarat Chandarlapaty, Tari A King, Jorge S Reis-Filho, Carlos L Arteaga, So Yeon Park, Franziska Michor and Kornelia Polyak (Oct. 2015). 'In situ single-cell analysis identifies heterogeneity for PIK3CA mutation and HER2 amplification in HER2-positive breast cancer'. en. In: *Nat. Genet.* 47.10, pp. 1212–1219.

Janiszewska, Michalina, Doris P Tabassum, Zafira Castaño, Simona Cristea, Kimiyo N Yamamoto, Natalie L Kingston, Katherine C Murphy, Shaokun Shu, Nicholas W Harper, Carlos Gil Del Alcazar, Maša Alečković, Muhammad B Ekram, Ofir Cohen, Minsuk Kwak, Yuanbo Qin, Tyler Laszewski, Adrienne Luoma, Andriy Marusyk,

Kai W Wucherpfennig, Nikhil Wagle, Rong Fan, Franziska Michor, Sandra S McAllister and Kornelia Polyak (July 2019). 'Subclonal cooperation drives metastasis by modulating local and systemic immune microenvironments'. en. In: *Nat. Cell Biol.* 21.7, pp. 879–888.

Jatoi, I, S G Hilsenbeck, G M Clark and C K Osborne (Aug. 1999). 'Significance of axillary lymph node metastasis in primary breast cancer'. en. In: *J. Clin. Oncol.* 17.8, pp. 2334–2340.

Javed, Asma and Aida Lteif (Feb. 2013). 'Development of the human breast'. en. In: *Semin. Plast. Surg.* 27.1, pp. 5–12.

Ji, Andrew L, Adam J Rubin, Kim Thrane, Sizun Jiang, David L Reynolds, Robin M Meyers, Margaret G Guo, Benson M George, Annelie Mollbrink, Joseph Bergenstråhle, Ludvig Larsson, Yunhao Bai, Bokai Zhu, Aparna Bhaduri, Jordan M Meyers, Xavier Rovira-Clavé, S Tyler Hollmig, Sumaira Z Aasi, Garry P Nolan, Joakim Lundeberg and Paul A Khavari (July 2020). 'Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma'. en. In: *Cell* 182.2, 497–514.e22.

Jones, Siân, Wei-Dong Chen, Giovanni Parmigiani, Frank Diehl, Niko Beerenwinkel, Tibor Antal, Arne Traulsen, Martin A Nowak, Christopher Siegel, Victor E Velculescu, Kenneth W Kinzler, Bert Vogelstein, Joseph Willis and Sanford D Markowitz (Mar. 2008). 'Comparative lesion sequencing provides insights into tumor evolution'. en. In: *Proc. Natl. Acad. Sci. U. S. A.* 105.11, pp. 4283–4288.

Kalluri, Raghu (Aug. 2016). 'The biology and function of fibroblasts in cancer'. en. In: *Nat. Rev. Cancer* 16.9, pp. 582–598.

Karlsson, Jenny, Anders Valind, Linda Holmquist Mengelbier, Sofia Bredin, Louise Cornmark, Caroline Jansson, Amina Wali, Johan Staaf, Björn Viklund, Ingrid Øra, Anna Börjesson, Torbjörn Backman, Noémie Braekeveldt, Bengt Sandstedt, Niklas Pal, Anders Isaksson, Barbara Gürtl Lackner, Tord Jonson, Daniel Bexell and David Gisselsson (July 2018). 'Four evolutionary trajectories underlie genetic intratumoral variation in childhood cancer'. en. In: *Nat. Genet.* 50.7, pp. 944–950.

Kather, Jakob Nikolas, Lara R Heij, Heike I Grabsch, Chiara Loeffler, Amelie Echle, Hannah Sophie Muti, Jeremias Krause, Jan M Niehues, Kai A J Sommer, Peter Bankhead, Loes F S Kooreman, Jefree J Schulte, Nicole A Cipriani, Roman D Buelow, Peter Boor, Nadi-Na Ortiz-Brüchle, Andrew M Hanby, Valerie Speirs, Sara Kochanny, Akash Patnaik, Andrew Srisuwananukorn, Hermann Brenner, Michael Hoffmeister, Piet A van den Brandt, Dirk Jäger, Christian Trautwein, Alexander T Pearson and Tom Luedde (Aug. 2020). 'Pan-cancer image-based detection of clinically actionable genetic alterations'. en. In: *Nat Cancer* 1.8, pp. 789–799.

Ke, Rongqin, Marco Mignardi, Alexandra Pacureanu, Jessica Svedlund, Johan Botling, Carolina Wählby and Mats Nilsson (Sept. 2013). 'In situ sequencing for RNA analysis in preserved tissue and cells'. en. In: *Nat. Methods* 10.9, pp. 857–860.

Keren, Leeat, Marc Bosse, Diana Marquez, Roshan Angoshtari, Samir Jain, Sushama Varma, Soo-Ryum Yang, Allison Kurian, David Van Valen, Robert West, Sean C Bendall and Michael Angelo (Sept. 2018). 'A Structured Tumor-Immune Microenvironment

in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging'. en. In: *Cell* 174.6, 1373–1387.e19.

Keren, Leeat, Marc Bosse, Steve Thompson, Tyler Risom, Kausalia Vijayaragavan, Erin McCaffrey, Diana Marquez, Roshan Angoshtari, Noah F Greenwald, Harris Fienberg, Jennifer Wang, Neeraja Kambham, David Kirkwood, Garry Nolan, Thomas J Montine, Stephen J Galli, Robert West, Sean C Bendall and Michael Angelo (Oct. 2019). 'MIBI-TOF: A multiplexed imaging platform relates cellular phenotypes and tissue structure'. en. In: *Sci Adv* 5.10, eaax5851.

Kiemen, Ashley, Alicia M Braxton, Mia P Grahn, Kyu Sang Han, Jaanvi Mahesh Babu, Rebecca Reichel, Falone Amoa, Seung-Mo Hong, Toby C Cornish, Elizabeth D Thompson, Laura D Wood, Ralph H Hruban, Pei-Hsun Wu and Denis Wirtz (Dec. 2020). 'In situ characterization of the 3D microanatomy of the pancreas and pancreatic cancer at single cell resolution'. en.

Kienast, Yvonne, Louisa von Baumgarten, Martin Fuhrmann, Wolfgang E F Klinkert, Roland Goldbrunner, Jochen Herms and Frank Winkler (Jan. 2010). 'Real-time imaging reveals the single steps of brain metastasis formation'. en. In: *Nat. Med.* 16.1, pp. 116–122.

Kim, Hee Kyung, Kyung Hee Park, Youjin Kim, Song Ee Park, Han Sang Lee, Sung Won Lim, Jang Ho Cho, Ji-Yeon Kim, Jeong Eon Lee, Jin Seok Ahn, Young-Hyuck Im, Jong Han Yu and Yeon Hee Park (Apr. 2019). 'Discordance of the PAM50 Intrinsic Subtypes Compared with Immunohistochemistry-Based Surrogate in Breast Cancer Patients: Potential Implication of Genomic Alterations of Discordance'. en. In: *Cancer Res. Treat.* 51.2, pp. 737–747.

Kingma, Diederik P and Jimmy Ba (Dec. 2014). 'Adam: A Method for Stochastic Optimization'. In: arXiv: 1412.6980 [cs.LG].

Kleshchevnikov, Vitalii, Artem Shmatko, Emma Dann, Alexander Aivazidis, Hamish W King, Tong Li, Rasa Elmentaite, Artem Lomakin, Veronika Kedlian, Adam Gayoso, Mika Sarkin Jain, Jun Sung Park, Lauma Ramona, Elizabeth Tuck, Anna Arutyunyan, Roser Vento-Tormo, Moritz Gerstung, Louisa James, Oliver Stegle and Omer Ali Bayraktar (Jan. 2022). 'Cell2location maps fine-grained cell types in spatial transcriptomics'. en. In: *Nat. Biotechnol.* 40.5, pp. 661–671.

Kohl, N E, C E Gee and F W Alt (Dec. 1984). 'Activated expression of the N-myc gene in human neuroblastomas and related tumors'. en. In: *Science* 226.4680, pp. 1335–1337.

Kole, Adam J, Henry S Park, Skyler B Johnson, Jacqueline R Kelly, Meena S Moran and Abhijit A Patel (July 2019). 'Overall survival is improved when DCIS accompanies invasive breast cancer'. en. In: *Sci. Rep.* 9.1, p. 9934.

Kozarewa, Iwanka, Zemin Ning, Michael A Quail, Mandy J Sanders, Matthew Berriman and Daniel J Turner (Apr. 2009). 'Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes'. en. In: *Nat. Methods* 6.4, pp. 291–295.

Krige, D G (1951). 'A statistical approach to some basic mine valuation problems on the Witwatersrand'. In: *J. South Afr. Inst. Min. Metall.*

Kucukelbir, Alp, Dustin Tran, Rajesh Ranganath, Andrew Gelman and David M Blei (Mar. 2016). 'Automatic Differentiation Variational Inference'. In: arXiv: 1603.00788 [stat.ML].

Kwon, Mi Jeong, Jinil Han, Ji Hyun Seo, Kyoung Song, Hae Min Jeong, Jong-Sun Choi, Yu Jin Kim, Seon-Heui Lee, Yoon-La Choi and Young Kee Shin (Oct. 2015). 'CD24 Overexpression Is Associated with Poor Prognosis in Luminal A and Triple-Negative Breast Cancer'. en. In: *PLoS One* 10.10, e0139112.

Laks, Emma, Andrew McPherson, Hans Zahn, Daniel Lai, Adi Steif, Jazmine Brimhall, Justina Biele, Beixi Wang, Tehmina Masud, Jerome Ting, Diljot Grewal, Cydney Nielsen, Samantha Leung, Viktoria Bojilova, Maia Smith, Oleg Golovko, Steven Poon, Peter Eirew, Farhia Kabeer, Teresa Ruiz de Algara, So Ra Lee, M Jafar Taghiyar, Curtis Huebner, Jessica Ngo, Tim Chan, Spencer Vatrt-Watts, Pascale Walters, Nafis Abrar, Sophia Chan, Matt Wiens, Lauren Martin, R Wilder Scott, T Michael Underhill, Elizabeth Chavez, Christian Steidl, Daniel Da Costa, Yussanne Ma, Robin J N Coope, Richard Corbett, Stephen Pleasance, Richard Moore, Andrew J Mungall, Colin Mar, Fergus Cafferty, Karen Gelmon, Stephen Chia, CRUK IMAXT Grand Challenge Team, Marco A Marra, Carl Hansen, Sohrab P Shah and Samuel Aparicio (Nov. 2019). 'Clonal Decomposition and DNA Replication States Defined by Scaled Single-Cell Genome Sequencing'. en. In: *Cell* 179.5, 1207–1221.e22.

Lambert, Arthur W, Diwakar R Pattabiraman and Robert A Weinberg (Feb. 2017). 'Emerging Biological Principles of Metastasis'. en. In: *Cell* 168.4, pp. 670–691.

Lambrechts, Diether, Els Wauters, Bram Boeckx, Sara Aibar, David Nittner, Oliver Burton, Ayse Bassez, Herbert Decaluwé, Andreas Pircher, Kathleen Van den Eynde, Birgit Weynand, Erik Verbeken, Paul De Leyn, Adrian Liston, Johan Vansteenkiste, Peter Carmeliet, Stein Aerts and Bernard Thienpont (Aug. 2018). 'Phenotype molding of stromal cells in the lung tumor microenvironment'. en. In: *Nat. Med.* 24.8, pp. 1277–1289.

Lane, D P and L V Crawford (Mar. 1979). 'T antigen is bound to a host protein in SV40-transformed cells'. en. In: *Nature* 278.5701, pp. 261–263.

Lareau, Caleb A, Fabiana M Duarte, Jennifer G Chew, Vinay K Kartha, Zach D Burkett, Andrew S Kohlway, Dmitry Pokholok, Martin J Aryee, Frank J Steemers, Ronald Lebofsky and Jason D Buenrostro (Aug. 2019). 'Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility'. en. In: *Nat. Biotechnol.* 37.8, pp. 916–924.

Larsson, Chatarina, Ida Grundberg, Ola Söderberg and Mats Nilsson (May 2010). 'In situ detection and genotyping of individual mRNA molecules'. en. In: *Nat. Methods* 7.5, pp. 395–397.

Lawrence, Michael S, Petar Stojanov, Paz Polak, Gregory V Kryukov, Kristian Cibulskis, Andrey Sivachenko, Scott L Carter, Chip Stewart, Craig H Mermel, Steven A Roberts, Adam Kiezun, Peter S Hammerman, Aaron McKenna, Yotam Drier, Lihua Zou, Alex H Ramos, Trevor J Pugh, Nicolas Stransky, Elena Helman, Jaegil Kim, Carrie Sougnez, Lauren Ambrogio, Elizabeth Nickerson, Erica Shefler, Maria L Cortés, Daniel Auclair, Gordon Saksena, Douglas Voet, Michael Noble, Daniel DiCara, Pei Lin, Lee Lichten-

stein, David I Heiman, Timothy Fennell, Marcin Imielinski, Bryan Hernandez, Eran Hodis, Sylvan Baca, Austin M Dulak, Jens Lohr, Dan-Avi Landau, Catherine J Wu, Jorge Melendez-Zajgla, Alfredo Hidalgo-Miranda, Amnon Koren, Steven A McCarroll, Jaume Mora, Brian Crompton, Robert Onofrio, Melissa Parkin, Wendy Winckler, Kristin Ardlie, Stacey B Gabriel, Charles W M Roberts, Jaclyn A Biegel, Kimberly Stegmaier, Adam J Bass, Levi A Garraway, Matthew Meyerson, Todd R Golub, Dmitry A Gordenin, Shamil Sunyaev, Eric S Lander and Gad Getz (July 2013). 'Mutational heterogeneity in cancer and the search for new cancer-associated genes'. en. In: *Nature* 499.7457, pp. 214–218.

Lawson, Andrew R J, Federico Abascal, Tim H H Coorens, Yvette Hooks, Laura O'Neill, Calli Latimer, Keiran Raine, Mathijs A Sanders, Anne Y Warren, Krishnaa T A Mahbubani, Bethany Bareham, Timothy M Butler, Luke M R Harvey, Alex Cagan, Andrew Menzies, Luiza Moore, Alexandra J Colquhoun, William Turner, Benjamin Thomas, Vincent Gnanapragasam, Nicholas Williams, Doris M Rassl, Harald Vöhringer, Sonia Zumalave, Jyoti Nangalia, José M C Tubío, Moritz Gerstung, Kourosh Saeb-Parsy, Michael R Stratton, Peter J Campbell, Thomas J Mitchell and Iñigo Martincorena (Oct. 2020). 'Extensive heterogeneity in somatic mutation and selection in the human bladder'. en. In: *Science* 370.6512, pp. 75–82.

Lee, Hee Jin, In Ah Park, In Hye Song, Su-Jin Shin, Joo Young Kim, Jong Han Yu and Gyungyub Gong (May 2016). 'Tertiary lymphoid structures: prognostic significance and relationship with tumour-infiltrating lymphocytes in triple-negative breast cancer'. en. In: *J. Clin. Pathol.* 69.5, pp. 422–430.

Lee, Hower, Sergio Marco Salas, Daniel Gyllborg and Mats Nilsson (May 2022). 'Direct RNA targeted in situ sequencing for transcriptomic profiling in tissue'. en. In: *Sci. Rep.* 12.1, p. 7976.

Lee, Je Hyuk, Evan R Daugharthy, Jonathan Scheiman, Reza Kalhor, Thomas C Ferrante, Richard Terry, Brian M Turczyk, Joyce L Yang, Ho Suk Lee, John Aach, Kun Zhang and George M Church (Mar. 2015). 'Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues'. en. In: *Nat. Protoc.* 10.3, pp. 442–458.

Lee-Six, Henry, Nina Friesgaard Øbro, Mairi S Shepherd, Sebastian Grossmann, Kevin Dawson, Miriam Belmonte, Robert J Osborne, Brian J P Huntly, Inigo Martincorena, Elizabeth Anderson, Laura O'Neill, Michael R Stratton, Elisa Laurenti, Anthony R Green, David G Kent and Peter J Campbell (Sept. 2018). 'Population dynamics of normal human blood inferred from somatic mutations'. en. In: *Nature* 561.7724, pp. 473–478.

Lee-Six, Henry, Sigurgeir Olafsson, Peter Ellis, Robert J Osborne, Mathijs A Sanders, Luiza Moore, Nikitas Georgakopoulos, Franco Torrente, Ayesha Noorani, Martin Goddard, Philip Robinson, Tim H H Coorens, Laura O'Neill, Christopher Alder, Jingwei Wang, Rebecca C Fitzgerald, Matthias Zilbauer, Nicholas Coleman, Kourosh Saeb-Parsy, Inigo Martincorena, Peter J Campbell and Michael R Stratton (Oct. 2019). 'The landscape of

somatic mutation in normal colorectal epithelial cells'. en. In: *Nature* 574.7779, pp. 532–537.

Lendahl, Urban, Lars Muhl and Christer Betsholtz (June 2022). 'Identification, discrimination and heterogeneity of fibroblasts'. en. In: *Nat. Commun.* 13.1, p. 3409.

Lewis, Sabrina M, Marie-Liesse Asselin-Labat, Quan Nguyen, Jean Berthelet, Xiao Tan, Verena C Wimmer, Delphine Merino, Kelly L Rogers and Shalin H Naik (Sept. 2021). 'Spatial omics and multiplexed imaging to explore cancer biology'. en. In: *Nat. Methods* 18.9, pp. 997–1012.

Li, Haoyang, Juexiao Zhou, Zhongxiao Li, Siyuan Chen, Xingyu Liao, Bin Zhang, Ruochi Zhang, Yu Wang, Shiwei Sun and Xin Gao (Mar. 2023). 'A comprehensive benchmarking with practical guidelines for cellular deconvolution of spatial transcriptomics'. en. In: *Nat. Commun.* 14.1, p. 1548.

Li, Heng and Richard Durbin (July 2009). 'Fast and accurate short read alignment with Burrows-Wheeler transform'. en. In: *Bioinformatics* 25.14, pp. 1754–1760.

Li, Ruoyan, Lin Di, Jie Li, Wenyi Fan, Yachen Liu, Wenjia Guo, Weiling Liu, Lu Liu, Qiong Li, Liping Chen, Yamei Chen, Chuanwang Miao, Hongjin Liu, Yuqian Wang, Yuling Ma, Deshu Xu, Dongxin Lin, Yanyi Huang, Jianbin Wang, Fan Bai and Chen Wu (Sept. 2021). 'A body map of somatic mutagenesis in morphologically normal human tissues'. en. In: *Nature* 597.7876, pp. 398–403.

Li, Tong, David Horsfall, Daniela Basurto-Lozada, Kenny Roberts, Martin Prete, John E G Lawrence, Peng He, Elisabeth Tuck, Josh Moore, Shila Ghazanfar, Sarah Teichmann, Muzlifah Haniffa and Omer Ali Bayraktar (May 2023). 'WebAtlas pipeline for integrated single cell and spatial transcriptomic data'. en.

Li, Xian-Peng, Xiao-Yu Yang, Ewelina Biskup, Jiang Zhou, Hong-Liang Li, Yi-Feng Wu, Ming-Liang Chen and Feng Xu (Sept. 2015). 'Co-expression of CXCL8 and HIF-1α is associated with metastasis and poor prognosis in hepatocellular carcinoma'. en. In: *Oncotarget* 6.26, pp. 22880–22889.

Liang, Paul Pu, Amir Zadeh and Louis-Philippe Morency (Sept. 2022). 'Foundations and Trends in Multimodal Machine Learning: Principles, Challenges, and Open Questions'. In: arXiv: 2209.03430 [cs.LG].

Lieberman, Erez, Christoph Hauert and Martin A Nowak (Jan. 2005). 'Evolutionary dynamics on graphs'. en. In: *Nature* 433.7023, pp. 312–316.

Lin, J - R (2021). 'Multiplexed 3D atlas of state transitions and immune interactions in colorectal cancer'. In: *bioRxiv*.

Lin, Jia-Ren, Mohammad Fallahi-Sichani and Peter K Sorger (Sept. 2015). 'Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method'. en. In: *Nat. Commun.* 6, p. 8390.

Lin, Jia-Ren, Benjamin Izar, Shu Wang, Clarence Yapp, Shaolin Mei, Parin M Shah, Sandro Santagata and Peter K Sorger (July 2018). 'Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes'. en. In: *Elife* 7, e31657.

Lips, Esther H, Tapsi Kumar, Anargyros Megalios, Lindy L Visser, Michael Sheinman, Angelo Fortunato, Vandna Shah, Marlous Hoogstraat, Emi Sei, Diego Mallo, Maria Roman-Escorza, Ahmed A Ahmed, Mingchu Xu, Alexandra W van den Belt-Dusebout, Wim Brugman, Anna K Casasent, Karen Clements, Helen R Davies, Liping Fu, Anita Grigoriadis, Timothy M Hardman, Lorraine M King, Marielle Krete, Petra Kristel, Michiel de Maaker, Carlo C Maley, Jeffrey R Marks, Brian A Menegaz, Lennart Mulder, Frank Nieboer, Salpie Nowinski, Sarah Pinder, Jelmar Quist, Carolina Salinas-Souza, Michael Schaapveld, Marjanka K Schmidt, Abeer M Shaaban, Rana Shami, Mathini Sridharan, John Zhang, Hilary Stobart, Deborah Collyar, Serena Nik-Zainal, Lodewyk F A Wessels, E Shelley Hwang, Nicholas E Navin, P Andrew Futreal, Grand Challenge PRECISION consortium, Alastair M Thompson, Jelle Wesseling and Elinor J Sawyer (June 2022). 'Genomic analysis defines clonal relationships of ductal carcinoma in situ and recurrent invasive breast cancer'. en. In: *Nat. Genet.* 54.6, pp. 850–860.

Liu, Yang, Mingyu Yang, Yanxiang Deng, Graham Su, Archibald Enninful, Cindy C Guo, Toma Tebaldi, Di Zhang, Dongjoo Kim, Zhiliang Bai, Eileen Norris, Alisia Pan, Jiatong Li, Yang Xiao, Stephanie Halene and Rong Fan (Dec. 2020). 'High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue'. en. In: *Cell* 183.6, 1665–1681.e18.

Loeffler, Chiara Maria Lavinia, Nadina Ortiz Bruechle, Max Jung, Lancelot Seillier, Michael Rose, Narmin Ghaffari Laleh, Ruth Knuechel, Titus J Brinker, Christian Trautwein, Nadine T Gaisa and Jakob N Kather (Mar. 2022). 'Artificial Intelligence-based Detection of FGFR3 Mutational Status Directly from Routine Histology in Bladder Cancer: A Possible Preselection for Molecular Testing?' en. In: *Eur Urol Focus* 8.2, pp. 472–479.

Lohman, Gregory J S, Robert J Bauer, Nicole M Nichols, Laurie Mazzola, Joanna Bybee, Danielle Rivizzigno, Elizabeth Cantin and Thomas C Evans Jr (Jan. 2016). 'A high-throughput assay for the comprehensive profiling of DNA ligase fidelity'. en. In: *Nucleic Acids Res.* 44.2, e14.

Lohoff, T, S Ghazanfar, A Missarova, N Koulena, N Pierson, J A Griffiths, E S Bardot, C-H L Eng, R C V Tyser, R Argelaguet, C Guibentif, S Srinivas, J Briscoe, B D Simons, A-K Hadjantonakis, B Göttgens, W Reik, J Nichols, L Cai and J C Marioni (Jan. 2022). 'Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis'. en. In: *Nat. Biotechnol.* 40.1, pp. 74–85.

Lomakin, Artem, Jessica Svedlund, Carina Strell, Milana Gataric, Artem Shmatko, Gleb Rukhovich, Jun Sung Park, Young Seok Ju, Stefan Dentro, Vitalii Kleshchevnikov, Vasyl Vaskivskyi, Tong Li, Omer Ali Bayraktar, Sarah Pinder, Andrea L Richardson, Sandro Santagata, Peter J Campbell, Hege Russnes, Moritz Gerstung, Mats Nilsson and Lucy R Yates (Nov. 2022). 'Spatial genomics maps the structure, nature and evolution of cancer clones'. en. In: *Nature* 611.7936, pp. 594–602.

Lopes, Joao V, Jorge M Pacheco and David Dingli (Dec. 2007). 'Acquired hematopoietic stem-cell disorders and mammalian size'. en. In: *Blood* 110.12, pp. 4120–4122.

Lopez, Romain, Baoguo Li, Hadas Keren-Shaul, Pierre Boyeau, Merav Kedmi, David Pilzer, Adam Jelinski, Ido Yofe, Eyal David, Allon Wagner, Can Ergen, Yoseph Addadi, Ofra Golani, Franca Ronchese, Michael I Jordan, Ido Amit and Nir Yosef (Sept. 2022). 'DestVI identifies continuums of cell types in spatial transcriptomics data'. en. In: *Nat. Biotechnol.* 40.9, pp. 1360–1369.

Louis, David N, Hiroko Ohgaki, Otmar D Wiestler, Webster K Cavenee, Peter C Burger, Anne Jouvet, Bernd W Scheithauer and Paul Kleihues (Aug. 2007). 'The 2007 WHO classification of tumours of the central nervous system'. en. In: *Acta Neuropathol.* 114.2, pp. 97–109.

Lu, Steve, Julie E Stein, David L Rimm, Daphne W Wang, J Michael Bell, Douglas B Johnson, Jeffrey A Sosman, Kurt A Schalper, Robert A Anders, Hao Wang, Clifford Hoyt, Drew M Pardoll, Ludmila Danilova and Janis M Taube (Aug. 2019). 'Comparison of Biomarker Modalities for Predicting Response to PD-1/PD-L1 Checkpoint Blockade: A Systematic Review and Meta-analysis'. en. In: *JAMA Oncol* 5.8, pp. 1195–1204.

Lu, Tian, Cheen Euong Ang and Xiaowei Zhuang (Nov. 2022). 'Spatially resolved epigenomic profiling of single cells in complex tissues'. en. In: *Cell* 185.23, 4448–4464.e17.

Luijt, P A van, E A M Heijnsdijk, J Fracheboud, L I H Overbeek, M J M Broeders, J Wesseling, G J den Heeten and H J de Koning (May 2016). 'The distribution of ductal carcinoma in situ (DCIS) grade in 4232 women and its impact on overdiagnosis in breast cancer screening'. en. In: *Breast Cancer Res.* 18.1, p. 47.

Łuksza, Marta, Zachary M Sethna, Luis A Rojas, Jayon Lihm, Barbara Bravi, Yuval Elhanati, Kevin Soares, Masataka Amisaki, Anton Dobrin, David Hoyos, Pablo Guasp, Abderezak Zebboudj, Rebecca Yu, Adrienne Kaya Chandra, Theresa Waters, Zagaa Odgerel, Joanne Leung, Rajya Kappagantula, Alvin Makohon-Moore, Amber Johns, Anthony Gill, Mathieu Gigoux, Jedd Wolchok, Taha Merghoub, Michel Sadelain, Erin Patterson, Remi Monasson, Thierry Mora, Aleksandra M Walczak, Simona Cocco, Christine Iacobuzio-Donahue, Benjamin D Greenbaum and Vinod P Balachandran (June 2022). 'Neoantigen quality predicts immunoediting in survivors of pancreatic cancer'. en. In: *Nature* 606.7913, pp. 389–395.

Lyubimova, Anna, Shalev Itzkovitz, Jan Philipp Junker, Zi Peng Fan, Xuebing Wu and Alexander van Oudenaarden (Sept. 2013). 'Single-molecule mRNA detection and counting in mammalian tissue'. en. In: *Nat. Protoc.* 8.9, pp. 1743–1758.

Ma, Ying and Xiang Zhou (Sept. 2022). 'Spatially informed cell-type deconvolution for spatial transcriptomics'. en. In: *Nat. Biotechnol.* 40.9, pp. 1349–1359.

Macaulay, Iain C, Wilfried Haerty, Parveen Kumar, Yang I Li, Tim Xiaoming Hu, Mabel J Teng, Mubeen Goolam, Nathalie Saurat, Paul Coupland, Lesley M Shirley, Miriam Smith, Niels Van der Aa, Ruby Banerjee, Peter D Ellis, Michael A Quail, Harold P Swerdlow, Magdalena Zernicka-Goetz, Frederick J Livesey, Chris P Ponting and Thierry Voet (June 2015). 'G&T-seq: parallel sequencing of single-cell genomes and transcriptomes'. en. In: *Nat. Methods* 12.6, pp. 519–522.

Macias, Hector and Lindsay Hinck (2012). 'Mammary gland development'. en. In: *Wiley Interdiscip. Rev. Dev. Biol.* 1.4, pp. 533–557.

Maldegem, Febe van, Karishma Valand, Megan Cole, Harshil Patel, Mihaela Angelova, Sareena Rana, Emma Colliver, Katey Enfield, Nourdine Bah, Gavin Kelly, Victoria Siu Kwan Tsang, Edurne Mugarza, Christopher Moore, Philip Hobson, Dina Levi, Miriam Molina-Arcas, Charles Swanton and Julian Downward (Oct. 2021). 'Characterisation of tumour microenvironment remodelling following oncogene inhibition in preclinical studies with imaging mass cytometry'. en. In: *Nat. Commun.* 12.1, p. 5906.

Mamlouk, Soulafa, Liam Harold Childs, Daniela Aust, Daniel Heim, Friederike Melching, Cristiano Oliveira, Thomas Wolf, Pawel Durek, Dirk Schumacher, Hendrik Bläker, Moritz von Winterfeld, Bastian Gastl, Kerstin Möhr, Andrea Menne, Silke Zeugner, Torben Redmer, Dido Lenze, Sascha Tierling, Markus Möbs, Wilko Weichert, Gunnar Folprecht, Eric Blanc, Dieter Beule, Reinhold Schäfer, Markus Morkel, Frederick Klauschen, Ulf Leser and Christine Sers (Jan. 2017). 'DNA copy number changes define spatial patterns of heterogeneity in colorectal cancer'. en. In: *Nat. Commun.* 8, p. 14093.

Martincorena, Iñigo and Peter J Campbell (Sept. 2015). 'Somatic mutation in cancer and normal cells'. en. In: *Science* 349.6255, pp. 1483–1489.

Martincorena, Iñigo, Joanna C Fowler, Agnieszka Wabik, Andrew R J Lawson, Federico Abascal, Michael W J Hall, Alex Cagan, Kasumi Murai, Krishnaa Mahbubani, Michael R Stratton, Rebecca C Fitzgerald, Penny A Handford, Peter J Campbell, Kourosh Saeb-Parsy and Philip H Jones (Nov. 2018). 'Somatic mutant clones colonize the human esophagus with age'. en. In: *Science* 362.6417, pp. 911–917.

Martincorena, Iñigo, Keiran M Raine, Moritz Gerstung, Kevin J Dawson, Kerstin Haase, Peter Van Loo, Helen Davies, Michael R Stratton and Peter J Campbell (Nov. 2017). 'Universal Patterns of Selection in Cancer and Somatic Tissues'. en. In: *Cell* 171.5, 1029–1041.e21.

Martincorena, Iñigo, Amit Roshan, Moritz Gerstung, Peter Ellis, Peter Van Loo, Stuart McLaren, David C Wedge, Anthony Fullam, Ludmil B Alexandrov, Jose M Tubio, Lucy Stebbings, Andrew Menzies, Sara Widaa, Michael R Stratton, Philip H Jones and Peter J Campbell (May 2015). 'Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin'. en. In: *Science* 348.6237, pp. 880–886.

Martínez-Jiménez, Francisco, Ali Movasati, Sascha Brunner, Luan Nguyen, Peter Priestley, Edwin Cuppen and Arne Van Hoeck (June 2022). 'Pan-cancer whole genome comparison of primary and metastatic solid tumors'. en.

McCaffrey, Erin F, Michele Donato, Leeat Keren, Zhenghao Chen, Alea Delmastro, Megan B Fitzpatrick, Sanjana Gupta, Noah F Greenwald, Alex Baranski, William Graf, Rashmi Kumar, Marc Bosse, Christine Camacho Fullaway, Pratista K Ramdial, Erna Forgó, Vladimir Jojic, David Van Valen, Smriti Mehra, Shabaana A Khader, Sean C Bendall, Matt van de Rijn, Daniel Kalman, Deepak Kaushal, Robert L Hunter, Niaz Banaei, Adrie J C Steyn, Purvesh Khatri and Michael Angelo (Feb. 2022). 'The immunoregulatory landscape of human tuberculosis granulomas'. en. In: *Nat. Immunol.* 23.2, pp. 318–329.

McCart Reed, Amy E, Lauren Kalinowski, Peter T Simpson and Sunil R Lakhani (Jan. 2021). 'Invasive lobular carcinoma of the breast: the increasing importance of this special subtype'. en. In: *Breast Cancer Res.* 23.1, p. 6.

McGranahan, Nicholas, Francesco Favero, Elza C de Bruin, Nicolai Juul Birkbak, Zoltan Szallasi and Charles Swanton (Apr. 2015). 'Clonal status of actionable driver events and the timing of mutational processes in cancer evolution'. en. In: *Sci. Transl. Med.* 7.283, 283ra54.

McGranahan, Nicholas, Andrew J S Furness, Rachel Rosenthal, Sofie Ramskov, Rikke Lyngaa, Sunil Kumar Saini, Mariam Jamal-Hanjani, Gareth A Wilson, Nicolai J Birkbak, Crispin T Hiley, Thomas B K Watkins, Seema Shafi, Nirupa Murugaesu, Richard Mitter, Ayse U Akarca, Joseph Linares, Teresa Marafioti, Jake Y Henry, Eliezer M Van Allen, Diana Miao, Bastian Schilling, Dirk Schadendorf, Levi A Garraway, Vladimir Makarov, Naiyer A Rizvi, Alexandra Snyder, Matthew D Hellmann, Taha Merghoub, Jedd D Wolchok, Sachet A Shukla, Catherine J Wu, Karl S Peggs, Timothy A Chan, Sine R Hadrup, Sergio A Quezada and Charles Swanton (Mar. 2016). 'Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade'. en. In: *Science* 351.6280, pp. 1463–1469.

McGranahan, Nicholas, Rachel Rosenthal, Crispin T Hiley, Andrew J Rowan, Thomas B K Watkins, Gareth A Wilson, Nicolai J Birkbak, Selvaraju Veeriah, Peter Van Loo, Javier Herrero, Charles Swanton and TRACERx Consortium (Nov. 2017). 'Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution'. en. In: *Cell* 171.6, 1259–1271.e11.

McPherson, Andrew, Andrew Roth, Emma Laks, Tehmina Masud, Ali Bashashati, Allen W Zhang, Gavin Ha, Justina Biele, Damian Yap, Adrian Wan, Leah M Prentice, Jaswinder Khattra, Maia A Smith, Cydney B Nielsen, Sarah C Mullaly, Steve Kalloger, Anthony Karnezis, Karey Shumansky, Celia Siu, Jamie Rosner, Hector Li Chan, Julie Ho, Nataliya Melnyk, Janine Senz, Winnie Yang, Richard Moore, Andrew J Mungall, Marco A Marra, Alexandre Bouchard-Côté, C Blake Gilks, David G Huntsman, Jessica N McAlpine, Samuel Aparicio and Sohrab P Shah (July 2016). 'Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer'. en. In: *Nat. Genet.* 48.7, pp. 758–767.

Miller, B E, F R Miller and G H Heppner (July 1989). 'Therapeutic perturbation of the tumor ecosystem in reconstructed heterogeneous mouse mammary tumors'. en. In: *Cancer Res.* 49.14, pp. 3747–3753.

Missarova, Alsu, Jaison Jain, Andrew Butler, Shila Ghazanfar, Tim Stuart, Maigan Brusko, Clive Wasserfall, Harry Nick, Todd Brusko, Mark Atkinson, Rahul Satija and John C Marioni (Dec. 2021). 'geneBasis: an iterative approach for unsupervised selection of targeted gene panels from scRNA-seq'. en. In: *Genome Biol.* 22.1, p. 333.

Moldoveanu, Dan, Leeann Ramsay, Mathieu Lajoie, Luke Anderson-Trocme, Marine Lingrand, Diana Berry, Lucas J M Perus, Yuhong Wei, Cleber Moraes, Rached Alkallas, Shivshankari Rajkumar, Dongmei Zuo, Matthew Dankner, Eric Hongbo Xu, Nicholas R Bertos, Hamed S Najafabadi, Simon Gravel, Santiago Costantino, Martin J Richer, Amanda W Lund, Sonia V Del Rincon, Alan Spatz, Wilson H Miller Jr, Rahima Jamal, Réjean Lapointe, Anne-Marie Mes-Masson, Simon Turcotte, Kevin Petrecca, Sinziana Dumitra, Ari-Nareg Meguerditchian, Keith Richardson, Francine Tremblay, Beatrice

Wang, May Chergui, Marie-Christine Guiot, Kevin Watters, John Stagg, Daniela F Quail, Catalin Mihalcioiu, Sarkis Meterissian and Ian R Watson (Apr. 2022). 'Spatially mapping the immune landscape of melanoma using imaging mass cytometry'. en. In: *Sci Immunol* 7.70, eabi5072.

Moncada, Reuben, Dalia Barkley, Florian Wagner, Marta Chiodin, Joseph C Devlin, Maayan Baron, Cristina H Hajdu, Diane M Simeone and Itai Yanai (Mar. 2020). 'Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas'. en. In: *Nat. Biotechnol.* 38.3, pp. 333–342.

Moore, Luiza, Alex Cagan, Tim H H Coorens, Matthew D C Neville, Rashesh Sanghvi, Mathijs A Sanders, Thomas R W Oliver, Daniel Leongamornlert, Peter Ellis, Ayesha Noorani, Thomas J Mitchell, Timothy M Butler, Yvette Hooks, Anne Y Warren, Mette Jorgensen, Kevin J Dawson, Andrew Menzies, Laura O'Neill, Calli Latimer, Mabel Teng, Ruben van Boxtel, Christine A Iacobuzio-Donahue, Inigo Martincorena, Rakesh Heer, Peter J Campbell, Rebecca C Fitzgerald, Michael R Stratton and Raheleh Rahbari (Sept. 2021). 'The mutational landscape of human somatic and germline cells'. en. In: *Nature* 597.7876, pp. 381–386.

Moore, Luiza, Daniel Leongamornlert, Tim H H Coorens, Mathijs A Sanders, Peter Ellis, Stefan C Dentro, Kevin J Dawson, Tim Butler, Raheleh Rahbari, Thomas J Mitchell, Francesco Maura, Jyoti Nangalia, Patrick S Tarpey, Simon F Brunner, Henry Lee-Six, Yvette Hooks, Sarah Moody, Krishnaa T Mahbubani, Mercedes Jimenez-Linan, Jan J Brosens, Christine A Iacobuzio-Donahue, Inigo Martincorena, Kourosh Saeb-Parsy, Peter J Campbell and Michael R Stratton (Apr. 2020). 'The mutational landscape of normal human endometrial epithelium'. en. In: *Nature* 580.7805, pp. 640–646.

Morin, P J, A B Sparks, V Korinek, N Barker, H Clevers, B Vogelstein and K W Kinzler (Mar. 1997). 'Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC'. en. In: *Science* 275.5307, pp. 1787–1790.

Morrissy, A Sorana, Florence M G Cavalli, Marc Remke, Vijay Ramaswamy, David J H Shih, Borja L Holgado, Hamza Farooq, Laura K Donovan, Livia Garzia, Sameer Agnihotri, Erin N Kiehna, Eloi Mercier, Chelsea Mayoh, Simon Papillon-Cavanagh, Hamid Nikbakht, Tenzin Gayden, Jonathon Torchia, Daniel Picard, Diana M Merino, Maria Vladoiu, Betty Luu, Xiaochong Wu, Craig Daniels, Stuart Horswell, Yuan Yao Thompson, Volker Hovestadt, Paul A Northcott, David T W Jones, John Peacock, Xin Wang, Stephen C Mack, Jüri Reimand, Steffen Albrecht, Adam M Fontebasso, Nina Thiessen, Yisu Li, Jacqueline E Schein, Darlene Lee, Rebecca Carlsen, Michael Mayo, Kane Tse, Angela Tam, Noreen Dhalla, Adrian Ally, Eric Chuah, Young Cheng, Patrick Plettner, Haiyan I Li, Richard D Corbett, Tina Wong, William Long, James Loukides, Pawel Buczkowicz, Cynthia E Hawkins, Uri Tabori, Brian R Rood, John S Myseros, Roger J Packer, Andrey Korshunov, Peter Lichter, Marcel Kool, Stefan M Pfister, Ulrich Schüller, Peter Dirks, Annie Huang, Eric Bouffet, James T Rutka, Gary D Bader, Charles Swanton, Yusanne Ma, Richard A Moore, Andrew J Mungall, Jacek Majewski, Steven J M Jones, Sunit Das, David Malkin, Nada Jabado, Marco A Marra and Michael

D Taylor (May 2017). 'Spatial heterogeneity in medulloblastoma'. en. In: *Nat. Genet.* 49.5, pp. 780–788.

Mund, Andreas, Andreas-David Brunner and Matthias Mann (June 2022). 'Unbiased spatial proteomics with single-cell resolution in tissues'. en. In: *Mol. Cell* 82.12, pp. 2335–2349.

Mund, Andreas, Fabian Coscia, András Kriston, Réka Hollandi, Ferenc Kovács, Andreas-David Brunner, Ede Migh, Lisa Schweizer, Alberto Santos, Michael Bzorek, Soraya Naimy, Lise Mette Rahbek-Gjerdrum, Beatrice Dyring-Andersen, Jutta Bulkescher, Claudia Lukas, Mark Adam Eckert, Ernst Lengyel, Christian Gnann, Emma Lundberg, Peter Horvath and Matthias Mann (Aug. 2022). 'Deep Visual Proteomics defines single-cell identity and heterogeneity'. en. In: *Nat. Biotechnol.* 40.8, pp. 1231–1240.

Murdoch, Craig, Munitta Muthana, Seth B Coffelt and Claire E Lewis (Aug. 2008). 'The role of myeloid cells in the promotion of tumour angiogenesis'. en. In: *Nat. Rev. Cancer* 8.8, pp. 618–631.

Nagasawa, Satoi, Yuta Kuze, Ichiro Maeda, Yasuyuki Kojima, Ai Motoyoshi, Tatsuya Onishi, Tsuguo Iwatani, Takamichi Yokoe, Junki Koike, Motohiro Chosokabe, Manabu Kubota, Hibiki Seino, Ayako Suzuki, Masahide Seki, Katsuya Tsuchihara, Eisuke Inoue, Koichiro Tsugawa, Tomohiko Ohta and Yutaka Suzuki (Apr. 2021). 'Genomic profiling reveals heterogeneous populations of ductal carcinoma in situ of the breast'. en. In: *Commun Biol* 4.1, p. 438.

Nam, Anna S, Ronan Chaligne and Dan A Landau (Jan. 2021). 'Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics'. en. In: *Nat. Rev. Genet.* 22.1, pp. 3–18.

Navin, Nicholas, Jude Kendall, Jennifer Troge, Peter Andrews, Linda Rodgers, Jeanne McIndoo, Kerry Cook, Asya Stepansky, Dan Levy, Diane Esposito, Lakshmi Muthuswamy, Alex Krasnitz, W Richard McCombie, James Hicks and Michael Wigler (Apr. 2011). 'Tumour evolution inferred by single-cell sequencing'. en. In: *Nature* 472.7341, pp. 90–94.

Navin, Nicholas, Alexander Krasnitz, Linda Rodgers, Kerry Cook, Jennifer Meth, Jude Kendall, Michael Riggs, Yvonne Eberling, Jennifer Troge, Vladimir Grubor, Dan Levy, Pär Lundin, Susanne Månér, Anders Zetterberg, James Hicks and Michael Wigler (Jan. 2010). 'Inferring tumor progression from genomic heterogeneity'. en. In: *Genome Res.* 20.1, pp. 68–80.

Nelson, M E, S G Riva and A Cvejic (Aug. 2022). 'SMaSH: a scalable, general marker gene identification framework for single-cell RNA-sequencing'. en. In: *BMC Bioinformatics* 23.1, p. 328.

Nguyen, Bastien, Christopher Fong, Anisha Luthra, Shaleigh A Smith, Renzo G DiNatale, Subhiksha Nandakumar, Henry Walch, Walid K Chatila, Ramyasree Madupuri, Ritika Kundra, Craig M Bielski, Brooke Mastrogiacomo, Mark T A Donoghue, Adrienne Boire, Sarat Chandarlapaty, Karuna Ganesh, James J Harding, Christine A Iacobuzio-Donahue, Pedram Razavi, Ed Reznik, Charles M Rudin, Dmitriy Zamarin, Wassim Abida, Ghassan K Abou-Alfa, Carol Aghajanian, Andrea Cercek, Ping Chi, Darren

Feldman, Alan L Ho, Gopakumar Iyer, Yelena Y Janjigian, Michael Morris, Robert J Motzer, Eileen M O'Reilly, Michael A Postow, Nitya P Raj, Gregory J Riely, Mark E Robson, Jonathan E Rosenberg, Anton Safonov, Alexander N Shoushtari, William Tap, Min Yuen Teo, Anna M Varghese, Martin Voss, Rona Yaeger, Marjorie G Zauderer, Nadeem Abu-Rustum, Julio Garcia-Aguilar, Bernard Bochner, Abraham Hakimi, William R Jarnagin, David R Jones, Daniela Molena, Luc Morris, Eric Rios-Doria, Paul Russo, Samuel Singer, Vivian E Strong, Debyani Chakravarty, Lora H Ellenson, Anuradha Gopalan, Jorge S Reis-Filho, Britta Weigelt, Marc Ladanyi, Mithat Gonen, Sohrab P Shah, Joan Massague, Jianjiong Gao, Ahmet Zehir, Michael F Berger, David B Solit, Samuel F Bakhoum, Francisco Sanchez-Vega and Nikolaus Schultz (Feb. 2022). 'Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients'. en. In: *Cell* 185.3, 563–575.e11.

Nguyen, Quy H, Nicholas Pervolarakis, Kerrigan Blake, Dennis Ma, Ryan Tevia Davis, Nathan James, Anh T Phung, Elizabeth Willey, Raj Kumar, Eric Jabart, Ian Driver, Jason Rock, Andrei Goga, Seema A Khan, Devon A Lawson, Zena Werb and Kai Kessenbrock (May 2018). 'Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity'. en. In: *Nat. Commun.* 9.1, p. 2028.

Ni, Zijian, Aman Prasad, Shuyang Chen, Richard B Halberg, Lisa M Arkin, Beth A Drolet, Michael A Newton and Christina Kendziorski (May 2022). 'SpotClean adjusts for spot swapping in spatial transcriptomics data'. en. In: *Nat. Commun.* 13.1, p. 2971.

Nichterwitz, Susanne, Geng Chen, Julio Aguila Benitez, Marlene Yilmaz, Helena Storvall, Ming Cao, Rickard Sandberg, Qiaolin Deng and Eva Hedlund (July 2016). 'Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling'. en. In: *Nat. Commun.* 7, p. 12139.

Nieto, Paula, Marc Elosua-Bayes, Juan L Trincado, Domenica Marchese, Ramon Massoni-Badosa, Maria Salvany, Ana Henriques, Juan Nieto, Sergio Aguilar-Fernández, Elisabetta Mereu, Catia Moutinho, Sara Ruiz, Patricia Lorden, Vanessa T Chin, Dominik Kaczorowski, Chia-Ling Chan, Richard Gallagher, Angela Chou, Ester Planas-Rigol, Carlota Rubio-Perez, Ivo Gut, Josep M Piulats, Joan Seoane, Joseph E Powell, Eduard Batlle and Holger Heyn (Oct. 2021). 'A single-cell tumor immune atlas for precision oncology'. en. In: *Genome Res.* 31.10, pp. 1913–1926.

Nik-Zainal, Serena, Ludmil B Alexandrov, David C Wedge, Peter Van Loo, Christopher D Greenman, Keiran Raine, David Jones, Jonathan Hinton, John Marshall, Lucy A Stebbings, Andrew Menzies, Sancha Martin, Kenric Leung, Lina Chen, Catherine Leroy, Manasa Ramakrishna, Richard Rance, King Wai Lau, Laura J Mudie, Ignacio Varela, David J McBride, Graham R Bignell, Susanna L Cooke, Adam Shlien, John Gamble, Ian Whitmore, Mark Maddison, Patrick S Tarpey, Helen R Davies, Elli Papaemmanuil, Philip J Stephens, Stuart McLaren, Adam P Butler, Jon W Teague, Göran Jönsson, Judy E Garber, Daniel Silver, Penelope Miron, Aquila Fatima, Sandrine Boyault, Anita Langerød, Andrew Tutt, John W M Martens, Samuel A J, Åke Borg, Anne Vincent Salomon, Gilles Thomas, Anne-Lise Børresen-Dale, Andrea L Richardson, Michael S Neuberger, P Andrew Futreal, Peter J Campbell and Michael R Stratton

(May 2012). 'Mutational Processes Molding the Genomes of 21 Breast Cancers'. en. In: *Cell* 149.5, pp. 979–993.

Nik-Zainal, Serena, Helen Davies, Johan Staaf, Manasa Ramakrishna, Dominik Glodzik, Xueqing Zou, Inigo Martincorena, Ludmil B Alexandrov, Sancha Martin, David C Wedge, Peter Van Loo, Young Seok Ju, Marcel Smid, Arie B Brinkman, Sandro Morganella, Miriam R Aure, Ole Christian Lingjærde, Anita Langerød, Markus Ringnér, Sung-Min Ahn, Sandrine Boyault, Jane E Brock, Annegien Broeks, Adam Butler, Christine Desmedt, Luc Dirix, Serge Dronov, Aquila Fatima, John A Foekens, Moritz Gerstung, Gerrit K J Hooijer, Se Jin Jang, David R Jones, Hyung-Yong Kim, Tari A King, Savitri Krishnamurthy, Hee Jin Lee, Jeong-Yeon Lee, Yilong Li, Stuart McLaren, Andrew Menzies, Ville Mustonen, Sarah O'Meara, Iris Pauporté, Xavier Pivot, Colin A Purdie, Keiran Raine, Kamna Ramakrishnan, F Germán Rodríguez-González, Gilles Romieu, Anieta M Sieuwerts, Peter T Simpson, Rebecca Shepherd, Lucy Stebbings, Olafur A Stefansson, Jon Teague, Stefania Tommasi, Isabelle Treilleux, Gert G Van den Eynden, Peter Vermeulen, Anne Vincent-Salomon, Lucy Yates, Carlos Caldas, Laura van't Veer, Andrew Tutt, Stian Knappskog, Benita Kiat Tee Tan, Jos Jonkers, Åke Borg, Naoto T Ueno, Christos Sotiriou, Alain Viari, P Andrew Futreal, Peter J Campbell, Paul N Span, Steven Van Laere, Sunil R Lakhani, Jorunn E Eyfjord, Alastair M Thompson, Ewan Birney, Hendrik G Stunnenberg, Marc J van de Vijver, John W M Martens, Anne-Lise Børresen-Dale, Andrea L Richardson, Gu Kong, Gilles Thomas and Michael R Stratton (June 2016). 'Landscape of somatic mutations in 560 breast cancer whole-genome sequences'. en. In: *Nature* 534.7605, pp. 47–54.

Nik-Zainal, Serena, Peter Van Loo, David C Wedge, Ludmil B Alexandrov, Christopher D Greenman, King Wai Lau, Keiran Raine, David Jones, John Marshall, Manasa Ramakrishna, Adam Shlien, Susanna L Cooke, Jonathan Hinton, Andrew Menzies, Lucy A Stebbings, Catherine Leroy, Mingming Jia, Richard Rance, Laura J Mudie, Stephen J Gamble, Philip J Stephens, Stuart McLaren, Patrick S Tarpey, Elli Papaemmanuil, Helen R Davies, Ignacio Varela, David J McBride, Graham R Bignell, Kenric Leung, Adam P Butler, Jon W Teague, Sancha Martin, Goran Jönsson, Odette Mariani, Sandrine Boyault, Penelope Miron, Aquila Fatima, Anita Langerød, Samuel A J R Aparicio, Andrew Tutt, Anieta M Sieuwerts, Åke Borg, Gilles Thomas, Anne Vincent Salomon, Andrea L Richardson, Anne-Lise Børresen-Dale, P Andrew Futreal, Michael R Stratton, Peter J Campbell and Breast Cancer Working Group of the International Cancer Genome Consortium (May 2012). 'The life history of 21 breast cancers'. en. In: *Cell* 149.5, pp. 994–1007.

Nirmal, Ajit J, Zoltan Maliga, Tuulia Vallius, Brian Quattrochi, Alyce A Chen, Connor A Jacobson, Roxanne J Pelletier, Clarence Yapp, Raquel Arias-Camison, Yu-An Chen, Christine G Lian, George F Murphy, Sandro Santagata and Peter K Sorger (June 2022). 'The Spatial Landscape of Progression and Immunoediting in Primary Melanoma at Single-Cell Resolution'. en. In: *Cancer Discovery* 12.6, pp. 1518–1541.

Nishimura, Tomomi, Nobuyuki Kakiuchi, Kenichi Yoshida, Takaki Sakurai, Tatsuki R Kataoka, Eiji Kondoh, Yoshitsugu Chigusa, Masahiko Kawai, Morio Sawada, Takuya

Inoue, Yasuhide Takeuchi, Hirona Maeda, Satoko Baba, Yusuke Shiozawa, Ryunosuke Saiki, Masahiro M Nakagawa, Yasuhito Nannya, Yotaro Ochi, Tomonori Hirano, Tomoe Nakagawa, Yukiko Inagaki-Kawata, Kosuke Aoki, Masahiro Hirata, Kosaku Nanki, Mami Matano, Megumu Saito, Eiji Suzuki, Masahiro Takada, Masahiro Kawashima, Kosuke Kawaguchi, Kenichi Chiba, Yuichi Shiraishi, Junko Takita, Satoru Miyano, Masaki Mandai, Toshiro Sato, Kengo Takeuchi, Hironori Haga, Masakazu Toi and Seishi Ogawa (Aug. 2023). 'Evolutionary histories of breast cancer and related clones'. en. In: *Nature* 620.7974, pp. 607–614.

Noble, Robert, Dominik Burri, Cécile Le Sueur, Jeanne Lemant, Yannick Viossat, Jakob Nikolas Kather and Niko Beerenwinkel (Feb. 2022). 'Spatial structure governs the mode of tumour evolution'. en. In: *Nat Ecol Evol* 6.2, pp. 207–217.

Noorani, Ayesha, Xiaodun Li, Martin Goddard, Jason Crawte, Ludmil B Alexandrov, Maria Secrier, Matthew D Eldridge, Lawrence Bower, Jamie Weaver, Pierre Lao-Sirieix, Inigo Martincorena, Irene Debiram-Beecham, Nicola Grehan, Shona MacRae, Shalini Malhotra, Ahmad Miremadi, Tabitha Thomas, Sarah Galbraith, Lorraine Petersen, Stephen D Preston, David Gilligan, Andrew Hindmarsh, Richard H Hardwick, Michael R Stratton, David C Wedge and Rebecca C Fitzgerald (Jan. 2020). 'Genomic evidence supports a clonal diaspora model for metastases of esophageal adenocarcinoma'. en. In: *Nat. Genet.* 52.1, pp. 74–83.

Nowak, Martin A, Franziska Michor and Yoh Iwasa (Dec. 2003). 'The linear process of somatic evolution'. en. In: *Proc. Natl. Acad. Sci. U. S. A.* 100.25, pp. 14966–14969.

Nowell, P C (Oct. 1976). 'The clonal evolution of tumor cell populations'. en. In: *Science* 194.4260, pp. 23–28.

Nowell, P C and D A Hungerford (July 1960). 'Chromosome studies on normal and leukemic human leukocytes'. en. In: *J. Natl. Cancer Inst.* 25, pp. 85–109.

Pagès, Franck, Bernhard Mlecnik, Florence Marliot, Gabriela Bindea, Fang-Shu Ou, Carlo Bifulco, Alessandro Lugli, Inti Zlobec, Tilman T Rau, Martin D Berger, Iris D Nagtegaal, Elisa Vink-Börger, Arndt Hartmann, Carol Geppert, Julie Kolwelter, Susanne Merkel, Robert Grützmann, Marc Van den Eynde, Anne Jouret-Mourin, Alex Kartheuser, Daniel Léonard, Christophe Remue, Julia Y Wang, Prashant Bavi, Michael H A Roehrl, Pamela S Ohashi, Linh T Nguyen, Seongjun Han, Heather L MacGregor, Sara Hafezi-Bakhtiari, Bradly G Wouters, Giuseppe V Masucci, Emilia K Andersson, Eva Zavadova, Michal Vocka, Jan Spacek, Lubos Petruzelka, Bohuslav Konopasek, Pavel Dundr, Helena Skalova, Kristyna Nemejcova, Gerardo Botti, Fabiana Tatangelo, Paolo Delrio, Gennaro Ciliberto, Michele Maio, Luigi Laghi, Fabio Grizzi, Tessa Fredriksen, Bénédicte Buttard, Mihaela Angelova, Angela Vasaturo, Pauline Maby, Sarah E Church, Helen K Angell, Lucie Lafontaine, Daniela Bruni, Carine El Sissy, Nacilla Haicheur, Amos Kirilovsky, Anne Berger, Christine Lagorce, Jeffrey P Meyers, Christopher Paustian, Zipei Feng, Carmen Ballesteros-Merino, Jeroen Dijkstra, Carlijn van de Water, Shannon van Lent-van Vliet, Nikki Knijn, Ana-Maria Mușină, Dragos-Viorel Scripcariu, Boryana Popivanova, Mingli Xu, Tomonobu Fujita, Shoichi Hazama, Nobuaki Suzuki, Hiroaki Nagano, Kiyotaka Okuno, Toshihiko Torigoe, Noriyuki Sato,

Tomohisa Furuhata, Ichiro Takemasa, Kyogo Itoh, Prabhu S Patel, Hemangini H Vora, Birva Shah, Jayendrakumar B Patel, Kruti N Rajvik, Shashank J Pandya, Shilin N Shukla, Yili Wang, Guanjun Zhang, Yutaka Kawakami, Francesco M Marincola, Paolo A Ascierto, Daniel J Sargent, Bernard A Fox and Jérôme Galon (May 2018). 'International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study'. en. In: *Lancet* 391.10135, pp. 2128–2139.

Paget, Stephen (Mar. 1889). 'THE DISTRIBUTION OF SECONDARY GROWTHS IN CANCER OF THE BREAST'. In: *Lancet* 133.3421, pp. 571–573.

Paik, Soonmyung, Steven Shak, Gong Tang, Chungyeul Kim, Joffre Baker, Maureen Cronin, Frederick L Baehner, Michael G Walker, Drew Watson, Taesung Park, William Hiller, Edwin R Fisher, D Lawrence Wickerham, John Bryant and Norman Wolmark (Dec. 2004). 'A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer'. en. In: *N. Engl. J. Med.* 351.27, pp. 2817–2826.

Pal, Bhupinder, Yunshun Chen, François Vaillant, Bianca D Capaldo, Rachel Joyce, Xiaoyu Song, Vanessa L Bryant, Jocelyn S Penington, Leon Di Stefano, Nina Tubau Ribera, Stephen Wilcox, Gregory B Mann, kConFab, Anthony T Papenfuss, Geoffrey J Lindeman, Gordon K Smyth and Jane E Visvader (June 2021). 'A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast'. en. In: *EMBO J.* 40.11, e107333.

Palla, Giovanni, David S Fischer, Aviv Regev and Fabian J Theis (Mar. 2022). 'Spatial components of molecular tissue biology'. en. In: *Nat. Biotechnol.* 40.3, pp. 308–318.

Park, Jeongbin, Wonyl Choi, Sebastian Tiesmeyer, Brian Long, Lars E Borm, Emma Garren, Thuc Nghi Nguyen, Bosiljka Tasic, Simone Codeluppi, Tobias Graf, Matthias Schlesner, Oliver Stegle, Roland Eils and Naveed Ishaque (June 2021). 'Cell segmentation-free inference of cell types from in situ transcriptomics data'. en. In: *Nat. Commun.* 12.1, p. 3545.

Parker, Joel S, Michael Mullins, Maggie C U Cheang, Samuel Leung, David Voduc, Tammi Vickery, Sherri Davies, Christiane Fauron, Xiaping He, Zhiyuan Hu, John F Quackenbush, Inge J Stijleman, Juan Palazzo, J S Marron, Andrew B Nobel, Elaine Mardis, Torsten O Nielsen, Matthew J Ellis, Charles M Perou and Philip S Bernard (Mar. 2009). 'Supervised risk predictor of breast cancer based on intrinsic subtypes'. en. In: *J. Clin. Oncol.* 27.8, pp. 1160–1167.

Paszke, Adam, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai and Soumith Chintala (Dec. 2019). 'PyTorch: An Imperative Style, High-Performance Deep Learning Library'. In: arXiv: 1912.01703 [cs.LG].

Peinado, Héctor, Haiying Zhang, Irina R Matei, Bruno Costa-Silva, Ayuko Hoshino, Goncalo Rodrigues, Bethan Psaila, Rosandra N Kaplan, Jacqueline F Bromberg, Yibin Kang, Mina J Bissell, Thomas R Cox, Amato J Giaccia, Janine T Erler, Sachie Hiratsuka, Cyrus

M Ghajar and David Lyden (May 2017). 'Pre-metastatic niches: organ-specific homes for metastases'. en. In: *Nat. Rev. Cancer* 17.5, pp. 302–317.

Pereira, Bernard, Suet-Feung Chin, Oscar M Rueda, Hans-Kristian Moen Vollan, Elena Provenzano, Helen A Bardwell, Michelle Pugh, Linda Jones, Roslin Russell, Stephen-John Sammut, Dana W Y Tsui, Bin Liu, Sarah-Jane Dawson, Jean Abraham, Helen Northen, John F Peden, Abhik Mukherjee, Gulisa Turashvili, Andrew R Green, Steve McKinney, Arusha Oloumi, Sohrab Shah, Nitzan Rosenfeld, Leigh Murphy, David R Bentley, Ian O Ellis, Arnie Purushotham, Sarah E Pinder, Anne-Lise Børresen-Dale, Helena M Earl, Paul D Pharoah, Mark T Ross, Samuel Aparicio and Carlos Caldas (May 2016). 'The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes'. en. In: *Nat. Commun.* 7, p. 11479.

Perkel, Jeffrey M (Aug. 2019). 'Starfish enterprise: finding RNA patterns in single cells'. en. In: *Nature* 572.7770, pp. 549–551.

Petukhov, Viktor, Rosalind J Xu, Ruslan A Soldatov, Paolo Cadinu, Konstantin Khodosevich, Jeffrey R Moffitt and Peter V Kharchenko (Mar. 2022). 'Cell segmentation in imaging-based spatial transcriptomics'. en. In: *Nat. Biotechnol.* 40.3, pp. 345–354.

Phan, Du, Neeraj Pradhan and Martin Jankowiak (Dec. 2019). 'Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro'. In: arXiv: 1912.11554 [stat.ML].

Philip, Mary and Andrea Schietinger (Apr. 2022). 'CD8+ T cell differentiation and dysfunction in cancer'. en. In: *Nat. Rev. Immunol.* 22.4, pp. 209–223.

Pielawski, Nicolas, Axel Andersson, Christophe Avenel, Andrea Behanova, Eduard Chelebian, Anna Klemm, Fredrik Nysjö, Leslie Solorzano and Carolina Wählby (May 2023). 'TissUUmaps 3: Improvements in interactive visualization, exploration, and quality assessment of large-scale spatial omics data'. en. In: *Heliyon* 9.5, e15306.

Pinder, Sarah E (May 2010). 'Ductal carcinoma in situ (DCIS): pathological features, differential diagnosis, prognostic factors and specimen evaluation'. en. In: *Mod. Pathol.* 23 Suppl 2, S8–13.

Polyak, Kornelia and Robert A Weinberg (Apr. 2009). 'Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits'. en. In: *Nat. Rev. Cancer* 9.4, pp. 265–273.

Pombo Antunes, Ana Rita, Isabelle Scheyltjens, Francesca Lodi, Julie Messiaen, Asier Antoranz, Johnny Duerinck, Daliya Kancheva, Liesbet Martens, Karen De Vlaminck, Hannah Van Hove, Signe Schmidt Kjølner Hansen, Francesca Maria Bosisio, Koen Van der Borght, Steven De Vleeschouwer, Raf Sciot, Luc Bouwens, Michiel Verfaillie, Niels Vandamme, Roosmarijn E Vandenbroucke, Olivier De Wever, Yvan Saeys, Martin Guilliams, Conny Gysemans, Bart Neyns, Frederik De Smet, Diether Lambrechts, Jo A Van Ginderachter and Kiavash Movahedi (Apr. 2021). 'Single-cell profiling of myeloid cells in glioblastoma across species and disease stage reveals macrophage competition and specialization'. en. In: *Nat. Neurosci.* 24.4, pp. 595–610.

Ponz de Leon, M, C Sacchetti, R Sassatelli, G Zanghieri, L Roncucci and A Scalmati (May 1990). 'Evidence for the existence of different types of large bowel tumor: suggestions

from the clinical data of a population-based registry'. en. In: *J. Surg. Oncol.* 44.1, pp. 35–43.

Prabhakaran, Sandhya (June 2022). 'Sparcle: assigning transcripts to cells in multiplexed images'. en. In: *Bioinform Adv* 2.1, vbac048.

Prater, Michael D, Valérie Petit, I Alasdair Russell, Rajshekhar R Giraddi, Mona Shehata, Suraj Menon, Reiner Schulte, Ivo Kalajzic, Nicola Rath, Michael F Olson, Daniel Metzger, Marisa M Faraldo, Marie-Ange Deugnier, Marina A Glukhova and John Stingl (Oct. 2014). 'Mammary stem cells have myoepithelial cell properties'. en. In: *Nat. Cell Biol.* 16.10, pp. 942–50, 1–7.

Priestley, Peter, Jonathan Baber, Martijn P Lolkema, Neeltje Steeghs, Ewart de Bruijn, Charles Shale, Korneel Duyvesteyn, Susan Haidari, Arne van Hoeck, Wendy Onstenk, Paul Roepman, Mircea Voda, Haiko J Bloemendal, Vivianne C G Tjan-Heijnen, Carla M L van Herpen, Mariette Labots, Petronella O Witteveen, Egbert F Smit, Stefan Sleijfer, Emile E Voest and Edwin Cuppen (Nov. 2019). 'Pan-cancer whole-genome analyses of metastatic solid tumours'. en. In: *Nature* 575.7781, pp. 210–216.

Puram, Sidharth V, Itay Tirosh, Anuraag S Parikh, Anoop P Patel, Keren Yizhak, Shawn Gillespie, Christopher Rodman, Christina L Luo, Edmund A Mroz, Kevin S Emerick, Daniel G Deschler, Mark A Varvares, Ravi Mylvaganam, Orit Rozenblatt-Rosen, James W Rocco, William C Faquin, Derrick T Lin, Aviv Regev and Bradley E Bernstein (Dec. 2017). 'Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer'. en. In: *Cell* 171.7, 1611–1624.e24.

Qi, Jingjing, Hongxiang Sun, Yao Zhang, Zhengting Wang, Zhenzhen Xun, Ziyi Li, Xinyu Ding, Rujuan Bao, Liwen Hong, Wenqing Jia, Fei Fang, Hongzhi Liu, Lei Chen, Jie Zhong, Duowu Zou, Lianxin Liu, Leng Han, Florent Ginhoux, Yingbin Liu, Youqiong Ye and Bing Su (Apr. 2022). 'Single-cell and spatial analysis reveal interaction of FAP+ fibroblasts and SPP1+ macrophages in colorectal cancer'. en. In: *Nat. Commun.* 13.1, p. 1742.

Qian, Xiaoyan, Kenneth D Harris, Thomas Hauling, Dimitris Nicoloutsopoulos, Ana B Muñoz-Manchado, Nathan Skene, Jens Hjerling-Leffler and Mats Nilsson (Jan. 2020). 'Probabilistic cell typing enables fine mapping of closely related cell types in situ'. en. In: *Nat. Methods* 17.1, pp. 101–106.

Quinn, Jeffrey J, Matthew G Jones, Ross A Okimoto, Shigeki Nanjo, Michelle M Chan, Nir Yosef, Trever G Bivona and Jonathan S Weissman (Feb. 2021). 'Single-cell lineages reveal the rates, routes, and drivers of metastasis in cancer xenografts'. en. In: *Science* 371.6532.

Rakha, Emad, Michael Toss and Cecily Quinn (Feb. 2022). 'Specific cell differentiation in breast cancer: a basis for histological classification'. en. In: *J. Clin. Pathol.* 75.2, pp. 76–84.

Rao, Anjali, Dalia Barkley, Gustavo S França and Itai Yanai (Aug. 2021). 'Exploring tissue architecture using spatial transcriptomics'. en. In: *Nature* 596.7871, pp. 211–220.

Raskov, Hans, Adile Orhan, Jan Pravsgaard Christensen and Ismail Gögenur (Jan. 2021). 'Cytotoxic CD8+ T cells in cancer and cancer immunotherapy'. en. In: *Br. J. Cancer* 124.2, pp. 359–367.

Ravi, Vidhya M, Nicolas Neidert, Paulina Will, Kevin Joseph, Julian P Maier, Jan Kückelhaus, Lea Vollmer, Jonathan M Goeldner, Simon P Behringer, Florian Scherer, Melanie Boerries, Marie Follo, Tobias Weiss, Daniel Delev, Julius Kernbach, Pamela Franco, Nils Schallner, Christine Dierks, Maria Stella Carro, Ulrich G Hofmann, Christian Fung, Roman Sankowski, Marco Prinz, Jürgen Beck, Henrike Salié, Bertram Bengsch, Oliver Schnell and Dieter Henrik Heiland (Feb. 2022). 'T-cell dysfunction in the glioblastoma microenvironment is mediated by myeloid cells releasing interleukin-10'. en. In: *Nat. Commun.* 13.1, p. 925.

Reiter, Johannes G, Marina Baretti, Jeffrey M Gerold, Alvin P Makohon-Moore, Adil Daud, Christine A Iacobuzio-Donahue, Nilofer S Azad, Kenneth W Kinzler, Martin A Nowak and Bert Vogelstein (Nov. 2019). 'An analysis of genetic heterogeneity in untreated cancers'. en. In: *Nat. Rev. Cancer* 19.11, pp. 639–650.

Retecki, Kuba, Milena Seweryn, Agnieszka Graczyk-Jarzynka and Malgorzata Bajor (Nov. 2021). 'The Immune Landscape of Breast Cancer: Strategies for Overcoming Immunotherapy Resistance'. en. In: *Cancers* 13.23.

Reticker-Flynn, Nathan E, Weiruo Zhang, Julia A Belk, Pamela A Basto, Nichole K Escalante, Genay O W Pilarowski, Alborz Bejnood, Maria M Martins, Justin A Kenkel, Ian L Linde, Sreya Bagchi, Robert Yuan, Serena Chang, Matthew H Spitzer, Yaron Carmi, Jiahan Cheng, Lorna L Tolentino, Okmi Choi, Nancy Wu, Christina S Kong, Andrew J Gentles, John B Sunwoo, Ansuman T Satpathy, Sylvia K Plevritis and Edgar G Engleman (May 2022). 'Lymph node colonization induces tumor-immune tolerance to promote distant metastasis'. en. In: *Cell* 185.11, 1924–1942.e23.

Rios, Anne C, Nai Yang Fu, Geoffrey J Lindeman and Jane E Visvader (Feb. 2014). 'In situ identification of bipotent stem cells in the mammary gland'. en. In: *Nature* 506.7488, pp. 322–327.

Risom, Tyler, David R Glass, Inna Averbukh, Candace C Liu, Alex Baranski, Adam Kagel, Erin F McCaffrey, Noah F Greenwald, Belén Rivero-Gutiérrez, Siri H Strand, Sushama Varma, Alex Kong, Leeat Keren, Sucheta Srivastava, Chunfang Zhu, Zumana Khair, Deborah J Veis, Katherine Deschryver, Sujay Vennam, Carlo Maley, E Shelley Hwang, Jeffrey R Marks, Sean C Bendall, Graham A Colditz, Robert B West and Michael Angelo (Jan. 2022). 'Transition to invasive breast cancer is associated with progressive changes in the structure and composition of tumor stroma'. en. In: *Cell* 185.2, 299–310.e18.

Roberts, Gareth O and Jeffrey S Rosenthal (Jan. 2004). 'General state space Markov chains and MCMC algorithms'. In: *Probab. Surv.* 1.none, pp. 20–71.

Rodriques, Samuel G, Robert R Stickels, Aleksandrina Goeva, Carly A Martin, Evan Murray, Charles R Vanderburg, Joshua Welch, Linlin M Chen, Fei Chen and Evan Z Macosko (Mar. 2019). 'Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution'. en. In: *Science* 363.6434, pp. 1463–1467.

Rooney, Michael S, Sachet A Shukla, Catherine J Wu, Gad Getz and Nir Hacohen (Jan. 2015). 'Molecular and genetic properties of tumors associated with local immune cytolytic activity'. en. In: *Cell* 160.1-2, pp. 48–61.

Rosenbluth, A W, M N Rosenbluth, A H Teller et al. (1953). 'Equation of state calculations by fast computing machines'. In: *The journal of.*

Ross-Innes, Caryn S, Jennifer Becq, Andrew Warren, R Keira Cheetham, Helen Northen, Maria O'Donovan, Shalini Malhotra, Massimiliano di Pietro, Sergii Ivakhno, Miao He, Jamie M J Weaver, Andy G Lynch, Zoya Kingsbury, Mark Ross, Sean Humphray, David Bentley and Rebecca C Fitzgerald (Sept. 2015). 'Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma'. en. In: *Nat. Genet.* 47.9, pp. 1038–1046.

Rozenblatt-Rosen, Orit, Aviv Regev, Philipp Oberdoerffer, Tal Nawy, Anna Hupalowska, Jennifer E Rood, Orr Ashenberg, Ethan Cerami, Robert J Coffey, Emek Demir, Li Ding, Edward D Esplin, James M Ford, Jeremy Goecks, Sharmistha Ghosh, Joe W Gray, Justin Guinney, Sean E Hanlon, Shannon K Hughes, E Shelley Hwang, Christine A Iacobuzio-Donahue, Judit Jané-Valbuena, Bruce E Johnson, Ken S Lau, Tracy Lively, Sarah A Mazzilli, Dana Pe'er, Sandro Santagata, Alex K Shalek, Denis Schapiro, Michael P Snyder, Peter K Sorger, Avrum E Spira, Sudhir Srivastava, Kai Tan, Robert B West, Elizabeth H Williams and Human Tumor Atlas Network (Apr. 2020). 'The Human Tumor Atlas Network: Charting Tumor Transitions across Space and Time at Single-Cell Resolution'. en. In: *Cell* 181.2, pp. 236–249.

Rue, Håvard, Sara Martino and Nicolas Chopin (Apr. 2009). 'Approximate Bayesian Inference for Latent Gaussian models by using Integrated Nested Laplace Approximations'. en. In: *J. R. Stat. Soc. Series B Stat. Methodol.* 71.2, pp. 319–392.

Sade-Feldman, Moshe, Yunxin J Jiao, Jonathan H Chen, Michael S Rooney, Michal Barzily-Rokni, Jean-Pierre Eliane, Stacey L Bjorgaard, Marc R Hammond, Hans Vitzthum, Shauna M Blackmon, Dennie T Frederick, Mehlika Hazar-Rethinam, Brandon A Nadres, Emily E Van Seventer, Sachet A Shukla, Keren Yizhak, John P Ray, Daniel Rosebrock, Dimitri Livitz, Viktor Adalsteinsson, Gad Getz, Lyn M Duncan, Bo Li, Ryan B Corcoran, Donald P Lawrence, Anat Stemmer-Rachamimov, Genevieve M Boland, Dan A Landau, Keith T Flaherty, Ryan J Sullivan and Nir Hacohen (Oct. 2017). 'Resistance to checkpoint blockade therapy through inactivation of antigen presentation'. en. In: *Nat. Commun.* 8.1, p. 1136.

Sahai, Erik, Igor Astsaturov, Edna Cukierman, David G DeNardo, Mikala Egeblad, Ronald M Evans, Douglas Fearon, Florian R Greten, Sunil R Hingorani, Tony Hunter, Richard O Hynes, Rakesh K Jain, Tobias Janowitz, Claus Jorgensen, Alec C Kimmelman, Mikhail G Kolonin, Robert G Maki, R Scott Powers, Ellen Puré, Daniel C Ramirez, Ruth Scherz-Shouval, Mara H Sherman, Sheila Stewart, Thea D Tlsty, David A Tuveson, Fiona M Watt, Valerie Weaver, Ashani T Weeraratna and Zena Werb (Mar. 2020). 'A framework for advancing our understanding of cancer-associated fibroblasts'. en. In: *Nat. Rev. Cancer* 20.3, pp. 174–186.

Saka, Sinem K, Yu Wang, Jocelyn Y Kishi, Allen Zhu, Yitian Zeng, Wenxin Xie, Koray Kirli, Clarence Yapp, Marcelo Cicconet, Brian J Beliveau, Sylvain W Lapan, Siyuan Yin, Millicent Lin, Edward S Boyden, Pascal S Kaeser, German Pihan, George M Church and Peng Yin (Sept. 2019). 'Immuno-SABER enables highly multiplexed and amplified protein imaging in tissues'. en. In: *Nat. Biotechnol.* 37.9, pp. 1080–1090.

Salehi, Sohrab, Farhia Kabeer, Nicholas Ceglia, Mirela Andronescu, Marc J Williams, Kieran R Campbell, Tehmina Masud, Beixi Wang, Justina Biele, Jazmine Brimhall, David Gee, Hakwoo Lee, Jerome Ting, Allen W Zhang, Hoa Tran, Ciara O'Flanagan, Fatemeh Dorri, Nicole Rusk, Teresa Ruiz de Algara, So Ra Lee, Brian Yu Chieh Cheng, Peter Eirew, Takako Kono, Jenifer Pham, Diljot Grewal, Daniel Lai, Richard Moore, Andrew J Mungall, Marco A Marra, IMAXT Consortium, Andrew McPherson, Alexandre Bouchard-Côté, Samuel Aparicio and Sohrab P Shah (July 2021). 'Clonal fitness inferred from time-series modelling of single-cell cancer genomes'. en. In: *Nature* 595.7868, pp. 585–590.

Salgado, R, C Denkert, S Demaria, N Sirtaine, F Klauschen, G Pruneri, S Wienert, G Van den Eynden, F L Baehner, F Penault-Llorca, E A Perez, E A Thompson, W F Symmans, A L Richardson, J Brock, C Criscitiello, H Bailey, M Ignatiadis, G Floris, J Sparano, Z Kos, T Nielsen, D L Rimm, K H Allison, J S Reis-Filho, S Loibl, C Sotiriou, G Viale, S Badve, S Adams, K Willard-Gallo, S Loi and International TILs Working Group 2014 (Feb. 2015). 'The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014'. en. In: *Ann. Oncol.* 26.2, pp. 259–271.

Salvatier, John, Thomas V Wiecki and Christopher Fonnesbeck (Apr. 2016). 'Probabilistic programming in Python using PyMC3'. en. In: *PeerJ Comput. Sci.* 2, e55.

Sautès-Fridman, Catherine, Florent Petitprez, Julien Calderaro and Wolf Herman Fridman (June 2019). 'Tertiary lymphoid structures in the era of cancer immunotherapy'. en. In: *Nat. Rev. Cancer* 19.6, pp. 307–325.

Savas, Peter, Balaji Virassamy, Chengzhong Ye, Agus Salim, Christopher P Mintoff, Franco Caramia, Roberto Salgado, David J Byrne, Zhi L Teo, Sathana Dushyanthen, Ann Byrne, Lironne Wein, Stephen J Luen, Catherine Poliness, Sophie S Nightingale, Anita S Skandarajah, David E Gyorki, Chantel M Thornton, Paul A Beavis, Stephen B Fox, Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer (kConFab), Phillip K Darcy, Terence P Speed, Laura K Mackay, Paul J Neeson and Sherene Loi (July 2018). 'Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis'. en. In: *Nat. Med.* 24.7, pp. 986–993.

Sawas, Tarek, Sarah Killcoyne, Prasad G Iyer, Kenneth K Wang, Thomas C Smyrk, John B Kisiel, Yi Qin, David A Ahlquist, Anil K Rustgi, Rui J Costa, Moritz Gerstung, Rebecca C Fitzgerald, David A Katzka and OCCAMS Consortium (Dec. 2018). 'Identification of Prognostic Phenotypes of Esophageal Adenocarcinoma in 2 Independent Cohorts'. en. In: *Gastroenterology* 155.6, 1720–1728.e4.

Schapiro, Denis, Hartland W Jackson, Swetha Raghuraman, Jana R Fischer, Vito R T Zanotelli, Daniel Schulz, Charlotte Giesen, Raúl Catena, Zsuzsanna Varga and Bernd Bodenmiller (Sept. 2017). 'histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data'. en. In: *Nat. Methods* 14.9, pp. 873–876.

Schapiro, Denis, Artem Sokolov, Clarence Yapp, Yu-An Chen, Jeremy L Muhlich, Joshua Hess, Allison L Creason, Ajit J Nirmal, Gregory J Baker, Maulik K Nariya, Jia-Ren Lin, Zoltan Maliga, Connor A Jacobson, Matthew W Hodgman, Juha Ruokonen, Samouil L Farhi, Domenic Abbondanza, Eliot T McKinley, Daniel Persson, Courtney Betts, Shamilene Sivagnanam, Aviv Regev, Jeremy Goecks, Robert J Coffey, Lisa M Coussens, Sandro Santagata and Peter K Sorger (Mar. 2022). 'MCMICRO: a scalable, modular image-processing pipeline for multiplexed tissue imaging'. en. In: *Nat. Methods* 19.3, pp. 311–315.

Schnitt, Stuart J and Laura C Collins (2013). *Biopsy interpretation of the breast.* en. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.

Schukken, Klaske M and Jason M Sheltzer (July 2022). 'Extensive protein dosage compensation in aneuploid human cancers'. en. In: *Genome Res.* 32.7, pp. 1254–1270.

Schumacher, Ton N and Daniela S Thommen (Jan. 2022). 'Tertiary lymphoid structures in cancer'. en. In: *Science* 375.6576, eabf9419.

Schürch, Christian M, Salil S Bhate, Graham L Barlow, Darci J Phillips, Luca Noti, Inti Zlobec, Pauline Chu, Sarah Black, Janos Demeter, David R McIlwain, Shigemi Kinoshita, Nikolay Samusik, Yury Goltsev and Garry P Nolan (Sept. 2020). 'Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front'. en. In: *Cell* 182.5, 1341–1359.e19.

Seferbekova, Zaira, Artem Lomakin, Lucy R Yates and Moritz Gerstung (May 2023). 'Spatial biology of cancer evolution'. en. In: *Nat. Rev. Genet.* 24.5, pp. 295–313.

Sender, Ron, Shai Fuchs and Ron Milo (Aug. 2016). 'Revised Estimates for the Number of Human and Bacteria Cells in the Body'. en. In: *PLoS Biol.* 14.8, e1002533.

Sender, Ron and Ron Milo (Jan. 2021). 'The distribution of cellular turnover in the human body'. en. In: *Nat. Med.* 27.1, pp. 45–48.

Sereesongsaeng, Naphannop, Sara H McDowell, James F Burrows, Christopher J Scott and Roberta E Burden (Dec. 2020). 'Cathepsin V suppresses GATA3 protein expression in luminal A breast cancer'. en. In: *Breast Cancer Res.* 22.1, p. 139.

Shah, Sohrab P, Ryan D Morin, Jaswinder Khattra, Leah Prentice, Trevor Pugh, Angela Burleigh, Allen Delaney, Karen Gelmon, Ryan Guliany, Janine Senz, Christian Steidl, Robert A Holt, Steven Jones, Mark Sun, Gillian Leung, Richard Moore, Tesa Severson, Greg A Taylor, Andrew E Teschendorff, Kane Tse, Gulisa Turashvili, Richard Varhol, René L Warren, Peter Watson, Yongjun Zhao, Carlos Caldas, David Huntsman, Martin Hirst, Marco A Marra and Samuel Aparicio (Oct. 2009). 'Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution'. en. In: *Nature* 461.7265, pp. 809–813.

Shah, Sohrab P, Andrew Roth, Rodrigo Goya, Arusha Oloumi, Gavin Ha, Yongjun Zhao, Gulisa Turashvili, Jiarui Ding, Kane Tse, Gholamreza Haffari, Ali Bashashati, Leah M

Prentice, Jaswinder Khattra, Angela Burleigh, Damian Yap, Virginie Bernard, Andrew McPherson, Karey Shumansky, Anamaria Crisan, Ryan Giuliany, Alireza Heravi-Moussavi, Jamie Rosner, Daniel Lai, Inanc Birol, Richard Varhol, Angela Tam, Noreen Dhalla, Thomas Zeng, Kevin Ma, Simon K Chan, Malachi Griffith, Annie Moradian, S-W Grace Cheng, Gregg B Morin, Peter Watson, Karen Gelmon, Stephen Chia, Suet-Feung Chin, Christina Curtis, Oscar M Rueda, Paul D Pharoah, Sambasivarao Damaraju, John Mackey, Kelly Hoon, Timothy Harkins, Vasisht Tadigotla, Mahvash Sigaroudinia, Philippe Gascard, Thea Tlsty, Joseph F Costello, Irmtraud M Meyer, Connie J Eaves, Wyeth W Wasserman, Steven Jones, David Huntsman, Martin Hirst, Carlos Caldas, Marco A Marra and Samuel Aparicio (Apr. 2012). 'The clonal and mutational evolution spectrum of primary triple-negative breast cancers'. en. In: *Nature* 486.7403, pp. 395–399.

Shao, Xin, Ning Lv, Jie Liao, Jinbo Long, Rui Xue, Ni Ai, Donghang Xu and Xiaohui Fan (Nov. 2019). 'Copy number variation is highly correlated with differential gene expression: a pan-cancer study'. en. In: *BMC Med. Genet.* 20.1, p. 175.

Sharonov, George V, Ekaterina O Serebrovskaya, Diana V Yuzhakova, Olga V Britanova and Dmitriy M Chudakov (May 2020). 'B cells, plasma cells and antibody repertoires in the tumour microenvironment'. en. In: *Nat. Rev. Immunol.* 20.5, pp. 294–307.

Shen, C Y, J C Yu, Y L Lo, C H Kuo, C T Yue, Y S Jou, C S Huang, J C Lung and C W Wu (July 2000). 'Genome-wide search for loss of heterozygosity using laser capture microdissected tissue of breast carcinoma: an implication for mutator phenotype and breast cancer pathogenesis'. en. In: *Cancer Res.* 60.14, pp. 3884–3892.

Shiozawa, Yusuke, Elisabeth A Pedersen, Aaron M Havens, Younghun Jung, Anjali Mishra, Jeena Joseph, Jin Koo Kim, Lalit R Patel, Chi Ying, Anne M Ziegler, Michael J Pienta, Junhui Song, Jingcheng Wang, Robert D Loberg, Paul H Krebsbach, Kenneth J Pienta and Russell S Taichman (Apr. 2011). 'Human prostate cancer metastases target the hematopoietic stem cell niche to establish footholds in mouse bone marrow'. en. In: *J. Clin. Invest.* 121.4, pp. 1298–1312.

Shmatko, Artem, Narmin Ghaffari Laleh, Moritz Gerstung and Jakob Nikolas Kather (Sept. 2022). 'Artificial intelligence in histopathology: enhancing cancer research and clinical oncology'. en. In: *Nat Cancer* 3.9, pp. 1026–1038.

Shukla, Sachet A, Michael S Rooney, Mohini Rajasagi, Grace Tiao, Philip M Dixon, Michael S Lawrence, Jonathan Stevens, William J Lane, Jamie L Dellagatta, Scott Steelman, Carrie Sougnez, Kristian Cibulskis, Adam Kiezun, Nir Hacohen, Vladimir Brusic, Catherine J Wu and Gad Getz (Nov. 2015). 'Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes'. en. In: *Nat. Biotechnol.* 33.11, pp. 1152–1158.

Sinha, Vidya C, Amanda L Rinkenbaugh, Mingchu Xu, Xinhui Zhou, Xiaomei Zhang, Sabrina Jeter-Jones, Jiansu Shao, Yuan Qi, John A Zebala, Dean Y Maeda, Florencia McAllister and Helen Piwnica-Worms (Aug. 2021). 'Single-cell evaluation reveals shifts in the tumor-immune niches that shape and maintain aggressive lesions in the breast'. en. In: *Nat. Commun.* 12.1, p. 5024.

Skibinski, A and C Kuperwasser (Oct. 2015). 'The origin of breast tumor heterogeneity'. en. In: *Oncogene* 34.42, pp. 5309–5316.

Slamon, D J, B Leyland-Jones, S Shak, H Fuchs, V Paton, A Bajamonde, T Fleming, W Eiermann, J Wolter, M Pegram, J Baselga and L Norton (Mar. 2001). 'Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2'. en. In: *N. Engl. J. Med.* 344.11, pp. 783–792.

Soldatov, Ruslan, Marketa Kaucka, Maria Eleni Kastriti, Julian Petersen, Tatiana Chontorotzea, Lukas Englmaier, Natalia Akkuratova, Yunshi Yang, Martin Häring, Viacheslav Dyachuk, Christoph Bock, Matthias Farlik, Michael L Piacentino, Franck Boismoreau, Markus M Hilscher, Chika Yokota, Xiaoyan Qian, Mats Nilsson, Marianne E Bronner, Laura Croci, Wen-Yu Hsiao, David A Guertin, Jean-Francois Brunet, Gian Giacomo Consalez, Patrik Ernfors, Kaj Fried, Peter V Kharchenko and Igor Adameyko (June 2019). 'Spatiotemporal structure of cell fate decisions in murine neural crest'. en. In: *Science* 364.6444.

Solin, Lawrence J, Robert Gray, Frederick L Baehner, Steven M Butler, Lorie L Hughes, Carl Yoshizawa, Diana B Cherbavaz, Steven Shak, David L Page, George W Sledge Jr, Nancy E Davidson, James N Ingle, Edith A Perez, William C Wood, Joseph A Sparano and Sunil Badve (May 2013). 'A multigene expression assay to predict local recurrence risk for ductal carcinoma in situ of the breast'. en. In: *J. Natl. Cancer Inst.* 105.10, pp. 701–710.

Somarelli, Jason A (Apr. 2021). 'The Hallmarks of Cancer as Ecologically Driven Phenotypes'. en. In: *Front Ecol Evol* 9.

Sondka, Zbyslaw, Sally Bamford, Charlotte G Cole, Sari A Ward, Ian Dunham and Simon A Forbes (Nov. 2018). 'The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers'. en. In: *Nat. Rev. Cancer* 18.11, pp. 696–705.

Sottoriva, Andrea, Haeyoun Kang, Zhicheng Ma, Trevor A Graham, Matthew P Salomon, Junsong Zhao, Paul Marjoram, Kimberly Siegmund, Michael F Press, Darryl Shibata and Christina Curtis (Mar. 2015). 'A Big Bang model of human colorectal tumor growth'. en. In: *Nat. Genet.* 47.3, pp. 209–216.

Spina, Elena and Pamela Cowin (June 2021). 'Embryonic mammary gland development'. en. In: *Semin. Cell Dev. Biol.* 114, pp. 83–92.

Stachler, Matthew D, Amaro Taylor-Weiner, Shouyong Peng, Aaron McKenna, Agoston T Agoston, Robert D Odze, Jon M Davison, Katie S Nason, Massimo Loda, Ignaty Leshchiner, Chip Stewart, Petar Stojanov, Sara Seepo, Michael S Lawrence, Daysha Ferrer-Torres, Jules Lin, Andrew C Chang, Stacey B Gabriel, Eric S Lander, David G Beer, Gad Getz, Scott L Carter and Adam J Bass (Sept. 2015). 'Paired exome analysis of Barrett's esophagus and adenocarcinoma'. en. In: *Nat. Genet.* 47.9, pp. 1047–1055.

Ståhl, Patrik L, Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, Michaela Asp, Jakub O Westholm, Mikael Huss, Annelie Mollbrink, Sten Linnarsson, Simone Codeluppi, Åke Borg, Fredrik Pontén, Paul Igor Costea, Pelin Sahlén, Jan Mulder, Olaf Bergmann, Joakim Lundeberg and Jo-

nas Frisén (July 2016). 'Visualization and analysis of gene expression in tissue sections by spatial transcriptomics'. en. In: *Science* 353.6294, pp. 78–82.

Stehelin, D, H E Varmus, J M Bishop and P K Vogt (Mar. 1976). 'DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA'. en. In: *Nature* 260.5547, pp. 170–173.

Stephens, Philip J, Patrick S Tarpey, Helen Davies, Peter Van Loo, Chris Greenman, David C Wedge, Serena Nik-Zainal, Sancha Martin, Ignacio Varela, Graham R Bignell, Lucy R Yates, Elli Papaemmanuil, David Beare, Adam Butler, Angela Cheverton, John Gamble, Jonathan Hinton, Mingming Jia, Alagu Jayakumar, David Jones, Calli Latimer, King Wai Lau, Stuart McLaren, David J McBride, Andrew Menzies, Laura Mudie, Keiran Raine, Roland Rad, Michael Spencer Chapman, Jon Teague, Douglas Easton, Anita Langerød, Oslo Breast Cancer Consortium (OSBREAC), Ming Ta Michael Lee, Chen-Yang Shen, Benita Tan Kiat Tee, Bernice Wong Huimin, Annegien Broeks, Ana Cristina Vargas, Gulisa Turashvili, John Martens, Aquila Fatima, Penelope Miron, Suet-Feung Chin, Gilles Thomas, Sandrine Boyault, Odette Mariani, Sunil R Lakhani, Marc van de Vijver, Laura van 't Veer, John Foekens, Christine Desmedt, Christos Sotiriou, Andrew Tutt, Carlos Caldas, Jorge S Reis-Filho, Samuel A J R Aparicio, Anne Vincent Salomon, Anne-Lise Børresen-Dale, Andrea L Richardson, Peter J Campbell, P Andrew Futreal and Michael R Stratton (May 2012). 'The landscape of cancer genes and mutational processes in breast cancer'. en. In: *Nature* 486.7403, pp. 400–404.

Stickels, Robert R, Evan Murray, Pawan Kumar, Jilong Li, Jamie L Marshall, Daniela J Di Bella, Paola Arlotta, Evan Z Macosko and Fei Chen (Mar. 2021). 'Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2'. en. In: *Nat. Biotechnol.* 39.3, pp. 313–319.

Stingl, J, C J Eaves, I Zandieh and J T Emerman (May 2001). 'Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue'. en. In: *Breast Cancer Res. Treat.* 67.2, pp. 93–109.

Stoeckius, Marlon, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K Chattopadhyay, Harold Swerdlow, Rahul Satija and Peter Smibert (Sept. 2017). 'Simultaneous epitope and transcriptome measurement in single cells'. en. In: *Nat. Methods* 14.9, pp. 865–868.

Stuart, Tim, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M Mauck 3rd, Yuhan Hao, Marlon Stoeckius, Peter Smibert and Rahul Satija (June 2019). 'Comprehensive Integration of Single-Cell Data'. en. In: *Cell* 177.7, 1888–1902.e21.

Su, Fei, Wei Zhang, Dalei Zhang, Yaqun Zhang, Cheng Pang, Yingying Huang, Miao Wang, Luwei Cui, Lei He, Jinsong Zhang, Lihui Zou, Junhua Zhang, Wenqinq Li, Lin Li, Jianyong Shao, Jie Ma, Fei Xiao and Ming Liu (Nov. 2018). 'Spatial Intratumor Genomic Heterogeneity within Localized Prostate Cancer Revealed by Single-nucleus Sequencing'. en. In: *Eur. Urol.* 74.5, pp. 551–559.

Su, Jun-Han, Pu Zheng, Seon S Kinrot, Bogdan Bintu and Xiaowei Zhuang (Sept. 2020). 'Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin'. en. In: *Cell* 182.6, 1641–1659.e26.

Sung, Hyuna, Jacques Ferlay, Rebecca L Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal and Freddie Bray (May 2021). 'Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries'. en. In: *CA Cancer J. Clin.* 71.3, pp. 209–249.

Svedlund, Jessica, Carina Strell, Xiaoyan Qian, Kilian J C Zilkens, Nicholas P Tobin, Jonas Bergh, Anieta M Sieuwerts and Mats Nilsson (Oct. 2019). 'Generation of in situ sequencing based OncoMaps to spatially resolve gene expression profiles of diagnostic and prognostic markers in breast cancer'. en. In: *EBioMedicine* 48, pp. 212–223.

Svensson, Valentine, Sarah A Teichmann and Oliver Stegle (May 2018). 'SpatialDE: identification of spatially variable genes'. en. In: *Nat. Methods* 15.5, pp. 343–346.

Szczerba, Barbara Maria, Francesc Castro-Giner, Marcus Vetter, Ilona Krol, Sofia Gkountela, Julia Landin, Manuel C Scheidmann, Cinzia Donato, Ramona Scherrer, Jochen Singer, Christian Beisel, Christian Kurzeder, Viola Heinzelmann-Schwarz, Christoph Rochlitz, Walter Paul Weber, Niko Beerenwinkel and Nicola Aceto (Feb. 2019). 'Neutrophils escort circulating tumour cells to enable cell cycle progression'. en. In: *Nature* 566.7745, pp. 553–557.

Tabár, László, Peter B Dean, Amy M-F Yen, Miklós Tarján, Sherry Y-H Chiu, Sam L-S Chen, Jean C-Y Fann and Tony H-H Chen (Feb. 2014). 'A Proposal to Unify the Classification of Breast and Prostate Cancers Based on the Anatomic Site of Cancer Origin and on Long-term Patient Outcome'. en. In: *Breast Cancer* 8, pp. 15–38.

Tammela, Tuomas, Francisco J Sanchez-Rivera, Naniye Malli Cetinbas, Katherine Wu, Nikhil S Joshi, Katja Helenius, Yoona Park, Roxana Azimi, Natanya R Kerper, R Alexander Wesselhoeft, Xin Gu, Leah Schmidt, Milton Cornwall-Brady, Ömer H Yilmaz, Wen Xue, Pekka Katajisto, Arjun Bhutkar and Tyler Jacks (May 2017). 'A Wnt-producing niche drives proliferative potential and progression in lung adenocarcinoma'. en. In: *Nature* 545.7654, pp. 355–359.

Tao, Luwei, Maaike P A van Bragt, Elizabeth Laudadio and Zhe Li (June 2014). 'Lineage tracing of mammary epithelial cells using cell-type-specific cre-expressing adenoviruses'. en. In: *Stem Cell Reports* 2.6, pp. 770–779.

Tape, Christopher J, Stephanie Ling, Maria Dimitriadi, Kelly M McMahon, Jonathan D Worboys, Hui Sun Leong, Ida C Norrie, Crispin J Miller, George Poulogiannis, Douglas A Lauffenburger and Claus Jørgensen (June 2016). 'Oncogenic KRAS Regulates Tumor Cell Signaling via Stromal Reciprocation'. en. In: *Cell* 165.7, p. 1818.

Tarabichi, Maxime, Adriana Salcedo, Amit G Deshwar, Máire Ni Leathlobhair, Jeff Wintersinger, David C Wedge, Peter Van Loo, Quaid D Morris and Paul C Boutros (Feb. 2021). 'A practical guide to cancer subclonal reconstruction from DNA sequencing'. en. In: *Nat. Methods* 18.2, pp. 144–155.

Taurin, Sebastien and Haifa Alkhalifa (Dec. 2020). 'Breast cancers, mammary stem cells, and cancer stem cells, characteristics, and hypotheses'. en. In: *Neoplasia* 22.12, pp. 663–678.

Tavernari, Daniele, Elena Battistello, Elie Dheilly, Aaron S Petruzzella, Marco Mina, Jessica Sordet-Dessimoz, Solange Peters, Thorsten Krueger, David Gfeller, Nicolo Riggi, Elisa Oricchio, Igor Letovanec and Giovanni Ciriello (June 2021). 'Nongenetic Evolution Drives Lung Adenocarcinoma Spatial Heterogeneity and Progression'. en. In: *Cancer Discov.* 11.6, pp. 1490–1507.

The Theano Development Team et al. (May 2016). 'Theano: A Python framework for fast computation of mathematical expressions'. In: arXiv: 1605.02688 [cs.SC].

Thomlinson, R H and L H Gray (Dec. 1955). 'The histological structure of some human lung cancers and the possible implications for radiotherapy'. en. In: *Br. J. Cancer* 9.4, pp. 539–549.

Thomson, J Z, A J Evans, S E Pinder, H C Burrell, A R Wilson and I O Ellis (July 2001). 'Growth pattern of ductal carcinoma in situ (DCIS): a retrospective analysis based on mammographic findings'. en. In: *Br. J. Cancer* 85.2, pp. 225–227.

Thorsson, Vésteinn, David L Gibbs, Scott D Brown, Denise Wolf, Dante S Bortone, Tai-Hsien Ou Yang, Eduard Porta-Pardo, Galen F Gao, Christopher L Plaisier, James A Eddy, Elad Ziv, Aedin C Culhane, Evan O Paull, I K Ashok Sivakumar, Andrew J Gentles, Raunaq Malhotra, Farshad Farshidfar, Antonio Colaprico, Joel S Parker, Lisle E Mose, Nam Sy Vo, Jianfang Liu, Yuexin Liu, Janet Rader, Varsha Dhankani, Sheila M Reynolds, Reanne Bowlby, Andrea Califano, Andrew D Cherniack, Dimitris Anastassiou, Davide Bedognetti, Younes Mokrab, Aaron M Newman, Arvind Rao, Ken Chen, Alexander Krasnitz, Hai Hu, Tathiane M Malta, Houtan Noushmehr, Chandra Sekhar Pedamallu, Susan Bullman, Akinyemi I Ojesina, Andrew Lamb, Wanding Zhou, Hui Shen, Toni K Choueiri, John N Weinstein, Justin Guinney, Joel Saltz, Robert A Holt, Charles S Rabkin, Cancer Genome Atlas Research Network, Alexander J Lazar, Jonathan S Serody, Elizabeth G Demicco, Mary L Disis, Benjamin G Vincent and Ilya Shmulevich (Aug. 2019). 'The Immune Landscape of Cancer'. en. In: *Immunity* 51.2, pp. 411–412.

Tirosh, Itay, Benjamin Izar, Sanjay M Prakadan, Marc H Wadsworth 2nd, Daniel Treacy, John J Trombetta, Asaf Rotem, Christopher Rodman, Christine Lian, George Murphy, Mohammad Fallahi-Sichani, Ken Dutton-Regester, Jia-Ren Lin, Ofir Cohen, Parin Shah, Diana Lu, Alex S Genshaft, Travis K Hughes, Carly G K Ziegler, Samuel W Kazer, Aleth Gaillard, Kellie E Kolb, Alexandra-Chloé Villani, Cory M Johannessen, Aleksandr Y Andreev, Eliezer M Van Allen, Monica Bertagnolli, Peter K Sorger, Ryan J Sullivan, Keith T Flaherty, Dennie T Frederick, Judit Jané-Valbuena, Charles H Yoon, Orit Rozenblatt-Rosen, Alex K Shalek, Aviv Regev and Levi A Garraway (Apr. 2016). 'Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq'. en. In: *Science* 352.6282, pp. 189–196.

Tkadlec, Josef, Andreas Pavlogiannis, Krishnendu Chatterjee and Martin A Nowak (Apr. 2019). 'Population structure determines the tradeoff between fixation probability and fixation time'. en. In: *Commun Biol* 2, p. 138.

Townes, F William and Barbara E Engelhardt (Feb. 2023). 'Nonnegative spatial factorization applied to spatial genomics'. en. In: *Nat. Methods* 20.2, pp. 229–238.

Tsai, Jeff H, Joana Liu Donaher, Danielle A Murphy, Sandra Chau and Jing Yang (Dec. 2012). 'Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis'. en. In: *Cancer Cell* 22.6, pp. 725–736.

Turajlic, Samra, Andrea Sottoriva, Trevor Graham and Charles Swanton (July 2019). 'Resolving genetic heterogeneity in cancer'. en. In: *Nat. Rev. Genet.* 20.7, pp. 404–416.

Ullah, Ikram, Govindasamy-Muralidharan Karthik, Amjad Alkodsi, Una Kjällquist, Gustav Stålhammar, John Lövrot, Nelson-Fuentes Martinez, Jens Lagergren, Sampsa Hautaniemi, Johan Hartman and Jonas Bergh (Apr. 2018). 'Evolutionary history of metastatic breast cancer reveals minimal seeding from axillary lymph nodes'. en. In: *J. Clin. Invest.* 128.4, pp. 1355–1370.

Valkenburg, Kenneth C, Amber E de Groot and Kenneth J Pienta (June 2018). 'Targeting the tumour stroma to improve cancer therapy'. en. In: *Nat. Rev. Clin. Oncol.* 15.6, pp. 366–381.

Van den Eynden, Jimmy, Alejandro Jiménez-Sánchez, Martin L Miller and Erik Larsson (Dec. 2019). 'Lack of detectable neoantigen depletion signals in the untreated cancer genome'. en. In: *Nat. Genet.* 51.12, pp. 1741–1748.

Van Keymeulen, Alexandra, Ana Sofia Rocha, Marielle Ousset, Benjamin Beck, Gaëlle Bouvencourt, Jason Rock, Neha Sharma, Sophie Dekoninck and Cédric Blanpain (Oct. 2011). 'Distinct stem cells contribute to mammary gland development and maintenance'. en. In: *Nature* 479.7372, pp. 189–193.

Velten, Britta, Jana M Braunger, Ricard Argelaguet, Damien Arnol, Jakob Wirbel, Danila Bredikhin, Georg Zeller and Oliver Stegle (Feb. 2022). 'Identifying temporal and spatial patterns of variation from multimodal data using MEFISTO'. en. In: *Nat. Methods* 19.2, pp. 179–186.

Venkataramani, Varun, Yvonne Yang, Marc Cicero Schubert, Ekin Reyhan, Svenja Kristin Tetzlaff, Niklas Wißmann, Michael Botz, Stella Judith Soyka, Carlo Antonio Beretta, Rangel Lyubomirov Pramatarov, Laura Fankhauser, Luciano Garofano, Alexander Freudenberg, Julia Wagner, Dimitar Ivanov Tanev, Miriam Ratliff, Ruifan Xie, Tobias Kessler, Dirk C Hoffmann, Ling Hai, Yvette Dörflinger, Simone Hoppe, Yahaya A Yabo, Anna Golebiewska, Simone P Niclou, Felix Sahm, Anna Lasorella, Martin Slowik, Leif Döring, Antonio Iavarone, Wolfgang Wick, Thomas Kuner and Frank Winkler (Aug. 2022). 'Glioblastoma hijacks neuronal mechanisms for brain invasion'. en. In: *Cell* 185.16, 2899–2917.e31.

Vennin, Claire, Pauline Mélénec, Romain Rouet, Max Nobis, Aurélie S Cazet, Kendelle J Murphy, David Herrmann, Daniel A Reed, Morghan C Lucas, Sean C Warren, Zehra Elgundi, Mark Pinese, Gabriella Kalna, Daniel Roden, Monisha Samuel, Anaiis Zaratzian, Shane T Grey, Andrew Da Silva, Wilfred Leung, Australian Pancreatic Genome

Initiative (APGI), Suresh Mathivanan, Yingxiao Wang, Anthony W Braithwaite, Daniel Christ, Ales Benda, Ashleigh Parkin, Phoebe A Phillips, John M Whitelock, Anthony J Gill, Owen J Sansom, David R Croucher, Benjamin L Parker, Marina Pajic, Jennifer P Morton, Thomas R Cox and Paul Timpson (Aug. 2019). 'CAF hierarchy driven by pancreatic cancer cell p53-status creates a pro-metastatic and chemoresistant environment via perlecan'. en. In: *Nat. Commun.* 10.1, p. 3637.

Venteicher, Andrew S, Itay Tirosh, Christine Hebert, Keren Yizhak, Cyril Neftel, Mariella G Filbin, Volker Hovestadt, Leah E Escalante, Mckenzie L Shaw, Christopher Rodman, Shawn M Gillespie, Danielle Dionne, Christina C Luo, Hiranmayi Ravichandran, Ravindra Mylvaganam, Christopher Mount, Maristela L Onozato, Brian V Nahed, Hiroaki Wakimoto, William T Curry, A John Iafrate, Miguel N Rivera, Matthew P Frosch, Todd R Golub, Priscilla K Brastianos, Gad Getz, Anoop P Patel, Michelle Monje, Daniel P Cahill, Orit Rozenblatt-Rosen, David N Louis, Bradley E Bernstein, Aviv Regev and Mario L Suvà (Mar. 2017). 'Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq'. en. In: *Science* 355.6332.

Vickovic, Sanja, Gökcen Eraslan, Fredrik Salmén, Johanna Klughammer, Linnea Stenbeck, Denis Schapiro, Tarmo Äijö, Richard Bonneau, Ludvig Bergenstråhle, José Fernandéz Navarro, Joshua Gould, Gabriel K Griffin, Åke Borg, Mostafa Ronaghi, Jonas Frisén, Joakim Lundeberg, Aviv Regev and Patrik L Ståhl (Oct. 2019). 'High-definition spatial transcriptomics for in situ tissue profiling'. en. In: *Nat. Methods* 16.10, pp. 987–990.

Waclaw, Bartlomiej, Ivana Bozic, Meredith E Pittman, Ralph H Hruban, Bert Vogelstein and Martin A Nowak (Sept. 2015). 'A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity'. en. In: *Nature* 525.7568, pp. 261–264.

Wagner, Johanna, Maria Anna Rapsomaniki, Stéphane Chevrier, Tobias Anzeneder, Claus Langwieder, August Dykgers, Martin Rees, Annette Ramaswamy, Simone Muenst, Savas Deniz Soysal, Andrea Jacobs, Jonas Windhager, Karina Silina, Maries van den Broek, Konstantin Johannes Dedes, Maria Rodríguez Martínez, Walter Paul Weber and Bernd Bodenmiller (May 2019). 'A Single-Cell Atlas of the Tumor and Immune Ecosystem of Human Breast Cancer'. en. In: *Cell* 177.5, 1330–1345.e18.

Waldman, Alex D, Jill M Fritz and Michael J Lenardo (Nov. 2020). 'A guide to cancer immunotherapy: from T cell basic science to clinical practice'. en. In: *Nat. Rev. Immunol.* 20.11, pp. 651–668.

Walt, Stéfan van der, Johannes L Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu and scikit-image contributors (June 2014). 'scikit-image: image processing in Python'. en. In: *PeerJ* 2, e453.

Wang, Fay, John Flanagan, Nan Su, Li-Chong Wang, Son Bui, Allissa Nielson, Xingyong Wu, Hong-Thuy Vo, Xiao-Jun Ma and Yuling Luo (Jan. 2012). 'RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues'. en. In: *J. Mol. Diagn.* 14.1, pp. 22–29.

Wang, Xiao, William E Allen, Matthew A Wright, Emily L Sylwestrak, Nikolay Samusik, Sam Vesuna, Kathryn Evans, Cindy Liu, Charu Ramakrishnan, Jia Liu, Garry P Nolan,

Felice-Alessio Bava and Karl Deisseroth (July 2018). 'Three-dimensional intact-tissue sequencing of single-cell transcriptional states'. en. In: *Science* 361.6400.

Wang, Xiao Qian, Esther Danenberg, Chiun-Sheng Huang, Daniel Egle, Maurizio Callari, Begoña Bermejo, Matteo Dugo, Claudio Zamagni, Marc Thill, Anton Anton, Stefania Zambelli, Stefania Russo, Eva Maria Ciruelos, Richard Greil, Balázs Győrffy, Vladimir Semiglazov, Marco Colleoni, Catherine M Kelly, Gabriella Mariani, Lucia Del Mastro, Olivia Biasi, Robert S Seitz, Pinuccia Valagussa, Giuseppe Viale, Luca Gianni, Giampaolo Bianchini and H Raza Ali (Sept. 2023). 'Spatial predictors of immunotherapy response in triple-negative breast cancer'. en. In: *Nature*.

Wang, Yong, Jill Waters, Marco L Leung, Anna Unruh, Whijae Roh, Xiuqing Shi, Ken Chen, Paul Scheet, Selina Vattathil, Han Liang, Asha Multani, Hong Zhang, Rui Zhao, Franziska Michor, Funda Meric-Bernstam and Nicholas E Navin (Aug. 2014). 'Clonal evolution in breast cancer revealed by single nucleus genome sequencing'. en. In: *Nature* 512.7513, pp. 155–160.

Watkins, Thomas B K, Emilia L Lim, Marina Petkovic, Sergi Elizalde, Nicolai J Birkbak, Gareth A Wilson, David A Moore, Eva Grönroos, Andrew Rowan, Sally M Dewhurst, Jonas Demeulemeester, Stefan C Dentro, Stuart Horswell, Lewis Au, Kerstin Haase, Mickael Escudero, Rachel Rosenthal, Maise Al Bakir, Hang Xu, Kevin Litchfield, Wei Ting Lu, Thanos P Mourikis, Michelle Dietzen, Lavinia Spain, George D Cresswell, Dhruva Biswas, Philippe Lamy, Iver Nordentoft, Katja Harbst, Francesc Castro-Giner, Lucy R Yates, Franco Caramia, Fanny Jaulin, Cécile Vicier, Ian P M Tomlinson, Priscilla K Brastianos, Raymond J Cho, Boris C Bastian, Lars Dyrskjøt, Göran B Jönsson, Peter Savas, Sherene Loi, Peter J Campbell, Fabrice Andre, Nicholas M Luscombe, Neeltje Steeghs, Vivianne C G Tjan-Heijnen, Zoltan Szallasi, Samra Turajlic, Mariam Jamal-Hanjani, Peter Van Loo, Samuel F Bakhoum, Roland F Schwarz, Nicholas McGranahan and Charles Swanton (Nov. 2020). 'Pervasive chromosomal instability and karyotype order in tumour evolution'. en. In: *Nature* 587.7832, pp. 126–132.

West, Jeffrey, Ryan O Schenck, Chandler Gatenbee, Mark Robertson-Tessi and Alexander R A Anderson (Apr. 2021). 'Normal tissue architecture determines the evolutionary course of cancer'. en. In: *Nat. Commun.* 12.1, p. 2060.

Wieland, Andreas, Mihir R Patel, Maria A Cardenas, Christiane S Eberhardt, William H Hudson, Rebecca C Obeng, Christopher C Griffith, Xu Wang, Zhuo G Chen, Haydn T Kissick, Nabil F Saba and Rafi Ahmed (Sept. 2021). 'Defining HPV-specific B cell responses in patients with head and neck cancer'. en. In: *Nature* 597.7875, pp. 274–278.

Williams, Jason B, Shuyin Li, Emily F Higgs, Alexandra Cabanov, Xiaozhong Wang, Haochu Huang and Thomas F Gajewski (Jan. 2020). 'Tumor heterogeneity and clonal cooperation influence the immune selection of IFN-γ-signaling mutant cancer cells'. en. In: *Nat. Commun.* 11.1, p. 602.

Williams, Marc J, Benjamin Werner, Chris P Barnes, Trevor A Graham and Andrea Sottoriva (Mar. 2016). 'Identification of neutral tumor evolution across cancer types'. en. In: *Nat. Genet.* 48.3, pp. 238–244.

Wolf, F Alexander, Philipp Angerer and Fabian J Theis (Feb. 2018). 'SCANPY: large-scale single-cell gene expression data analysis'. en. In: *Genome Biol.* 19.1, p. 15.

Wolf, Natalie K, Djem U Kissiov and David H Raulet (Feb. 2023). 'Roles of natural killer cells in immunity to cancer, and applications to immunotherapy'. en. In: *Nat. Rev. Immunol.* 23.2, pp. 90–105.

Woodcock, D J, E Riabchenko, S Taavitsainen, M Kankainen, G Gundem, D S Brewer, P Ellonen, M Lepistö, Y A Golubeva, A C Warner, T Tolonen, J Jasu, W B Isaacs, M R Emmert-Buck, M Nykter, T Visakorpi, G S Bova and D C Wedge (Oct. 2020). 'Prostate cancer evolution from multilineage primary to single lineage metastases with implications for liquid biopsy'. en. In: *Nat. Commun.* 11.1, p. 5070.

Wooster, R, S L Neuhausen, J Mangion, Y Quirk, D Ford, N Collins, K Nguyen, S Seal, T Tran and D Averill (Sept. 1994). 'Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13'. en. In: *Science* 265.5181, pp. 2088–2090.

Wu, Rui, Wenbo Guo, Xinyao Qiu, Shicheng Wang, Chengjun Sui, Qiuyu Lian, Jianmin Wu, Yiran Shan, Zhao Yang, Shuai Yang, Tong Wu, Kaiting Wang, Yanjing Zhu, Shan Wang, Changyi Liu, Yangqianwen Zhang, Bo Zheng, Zhixuan Li, Yani Zhang, Siyun Shen, Yan Zhao, Wenwen Wang, Jinxia Bao, Ji Hu, Xuan Wu, Xiaoqing Jiang, Hongyang Wang, Jin Gu and Lei Chen (Dec. 2021). 'Comprehensive analysis of spatial architecture in primary liver cancer'. en. In: *Sci Adv* 7.51, eabg3750.

Wu, Sunny Z, Ghamdan Al-Eryani, Daniel Lee Roden, Simon Junankar, Kate Harvey, Alma Andersson, Aatish Thennavan, Chenfei Wang, James R Torpy, Nenad Bartonicek, Taopeng Wang, Ludvig Larsson, Dominik Kaczorowski, Neil I Weisenfeld, Cedric R Uytingco, Jennifer G Chew, Zachary W Bent, Chia-Ling Chan, Vikkitharan Gnanasambandapillai, Charles-Antoine Dutertre, Laurence Gluch, Mun N Hui, Jane Beith, Andrew Parker, Elizabeth Robbins, Davendra Segara, Caroline Cooper, Cindy Mak, Belinda Chan, Sanjay Warrier, Florent Ginhoux, Ewan Millar, Joseph E Powell, Stephen R Williams, X Shirley Liu, Sandra O'Toole, Elgene Lim, Joakim Lundeberg, Charles M Perou and Alexander Swarbrick (Sept. 2021). 'A single-cell and spatially resolved atlas of human breast cancers'. en. In: *Nat. Genet.* 53.9, pp. 1334–1347.

Yates, Lucy R, Moritz Gerstung, Stian Knappskog, Christine Desmedt, Gunes Gundem, Peter Van Loo, Turid Aas, Ludmil B Alexandrov, Denis Larsimont, Helen Davies, Yilong Li, Young Seok Ju, Manasa Ramakrishna, Hans Kristian Haugland, Peer Kaare Lilleng, Serena Nik-Zainal, Stuart McLaren, Adam Butler, Sancha Martin, Dominic Glodzik, Andrew Menzies, Keiran Raine, Jonathan Hinton, David Jones, Laura J Mudie, Bing Jiang, Delphine Vincent, April Greene-Colozzi, Pierre-Yves Adnet, Aquila Fatima, Marion Maetens, Michail Ignatiadis, Michael R Stratton, Christos Sotiriou, Andrea L Richardson, Per Eystein Lønning, David C Wedge and Peter J Campbell (June 2015). 'Subclonal diversification of primary breast cancer revealed by multiregion sequencing'. en. In: *Nat. Med.* 21.7, pp. 751–759.

Yates, Lucy R, Stian Knappskog, David Wedge, James H R Farmery, Santiago Gonzalez, Inigo Martincorena, Ludmil B Alexandrov, Peter Van Loo, Hans Kristian Haugland, Peer Kaare Lilleng, Gunes Gundem, Moritz Gerstung, Elli Pappaemmanuil, Patrycja

Gazinska, Shriram G Bhosle, David Jones, Keiran Raine, Laura Mudie, Calli Latimer, Elinor Sawyer, Christine Desmedt, Christos Sotiriou, Michael R Stratton, Anieta M Sieuwerts, Andy G Lynch, John W Martens, Andrea L Richardson, Andrew Tutt, Per Eystein Lønning and Peter J Campbell (Aug. 2017). 'Genomic Evolution of Breast Cancer Metastasis and Relapse'. en. In: *Cancer Cell* 32.2, 169–184.e7.

Zaretsky, Jesse M, Angel Garcia-Diaz, Daniel S Shin, Helena Escuin-Ordinas, Willy Hugo, Siwen Hu-Lieskovan, Davis Y Torrejon, Gabriel Abril-Rodriguez, Salemiz Sandoval, Lucas Barthly, Justin Saco, Blanca Homet Moreno, Riccardo Mezzadra, Bartosz Chmielowski, Kathleen Ruchalski, I Peter Shintaku, Phillip J Sanchez, Cristina Puig-Saus, Grace Cherry, Elizabeth Seja, Xiangju Kong, Jia Pang, Beata Berent-Maoz, Begoña Comin-Anduix, Thomas G Graeber, Paul C Tumeh, Ton N M Schumacher, Roger S Lo and Antoni Ribas (Sept. 2016). 'Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma'. en. In: *N. Engl. J. Med.* 375.9, pp. 819–829.

Zeng, Qiqun, Iacovos P Michael, Peng Zhang, Sadegh Saghafinia, Graham Knott, Wei Jiao, Brian D McCabe, José A Galván, Hugh P C Robinson, Inti Zlobec, Giovanni Ciriello and Douglas Hanahan (Sept. 2019). 'Synaptic proximity enables NMDAR signalling to promote brain metastasis'. en. In: *Nature* 573.7775, pp. 526–531.

Zhang, Y (2022). 'Reference-based cell type matching of spatial transcriptomics data'. In: *bioRxiv*.

Zhao, Tongtong, Zachary D Chiang, Julia W Morriss, Lindsay M LaFave, Evan M Murray, Isabella Del Priore, Kevin Meli, Caleb A Lareau, Naeem M Nadaf, Jilong Li, Andrew S Earl, Evan Z Macosko, Tyler Jacks, Jason D Buenrostro and Fei Chen (Jan. 2022). 'Spatial genomics enables multi-modal study of clonal heterogeneity in tissues'. en. In: *Nature* 601.7891, pp. 85–91.

Zhou, Hengbo, Deepika Neelakantan and Heide L Ford (Apr. 2017). 'Clonal cooperativity in heterogenous cancers'. en. In: *Semin. Cell Dev. Biol.* 64, pp. 79–89.

Zhou, Huaqiang, Yi Hu, Rongzhen Luo, Yuanyuan Zhao, Hui Pan, Liyan Ji, Ting Zhou, Lanjun Zhang, Hao Long, Jianhua Fu, Zhesheng Wen, Siyu Wang, Xin Wang, Peng Lin, Haoxian Yang, Junye Wang, Mengmeng Song, Xin Yi, Ling Yang, Xuefang Xia, Yanfang Guan, Wenfeng Fang, Yunpeng Yang, Shaodong Hong, Yan Huang, Pansong Li, Yaxiong Zhang and Ningning Zhou (Sept. 2021). 'Multi-region exome sequencing reveals the intratumoral heterogeneity of surgically resected small cell lung cancer'. en. In: *Nat. Commun.* 12.1, p. 5431.