

 Open access • Posted Content • DOI:10.1101/2021.04.16.439912

## Spatial genomics maps the structure, character and evolution of cancer clones

— [Source link](#) 

Lomakin A, Lomakin A, Svedlund J, Svedlund J ...+22 more authors

**Institutions:** European Bioinformatics Institute, Wellcome Trust Sanger Institute, Science for Life Laboratory, Uppsala University ...+4 more institutions

**Published on:** 16 Apr 2021 - bioRxiv (Cold Spring Harbor Laboratory)

**Topics:** Genomics

Related papers:

- [Inference of clonal selection in cancer populations using single-cell sequencing data.](#)
- [Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes](#)
- [Revealing the inherent heterogeneity of human malignancies by variant consensus strategies coupled with cancer clonal analysis](#)
- [De novo spatiotemporal modelling of cell-type signatures identifies novel cell populations in the developmental human heart](#)
- [Pathway-based dissection of the genomic heterogeneity of cancer hallmarks' acquisition with SLAPenrich](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/spatial-genomics-maps-the-structure-character-and-evolution-rf3spaql7j>

# Spatial genomics maps the structure, character and evolution of cancer clones

Artem Lomakin<sup>1,2†</sup>, Jessica Svedlund<sup>3,4†</sup>, Carina Strell<sup>3,4</sup>, Milana Gataric<sup>1</sup>, Artem Shmatko<sup>1,2,5</sup>, Jun Sung Park<sup>1,2</sup>, Young Seok Ju<sup>6</sup>, Stefan Dentro<sup>1,2</sup>, Vitalii Kleshchevnikov<sup>2</sup>, Vasyl Vaskivskiy<sup>2</sup>, Tong Li<sup>2</sup>, Omer Ali Bayraktar<sup>2</sup>, Luiza Moore<sup>2</sup>, Sarah Pinder<sup>7,8</sup>, Andrea L Richardson<sup>9</sup>, Peter J Campbell<sup>2</sup>, Moritz Gerstung<sup>1,10\*</sup>, Mats Nilsson<sup>3\*</sup>, Lucy R Yates<sup>2\*</sup>

## Affiliations

<sup>1</sup>European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI); Hinxton, UK

<sup>2</sup>Wellcome Sanger Institute; Cambridge, UK, CB10 1SA

<sup>3</sup>Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University; 171 65 Solna, Sweden.

<sup>4</sup>Department of Immunology Genetics, and Pathology, Uppsala University; Uppsala, Sweden

<sup>5</sup>Moscow State University; Moscow, Russia

<sup>6</sup>Laboratory of Cancer Genomics, GSMSE, KAIST; Daejeon, Korea

<sup>7</sup>Guys and St Thomas' NHS Trust; London, UK

<sup>8</sup>Department of Pathology, King's College London; London, UK

<sup>9</sup>Department of Pathology, Johns Hopkins Medicine; Baltimore, MD, USA

<sup>10</sup>Genome Biology Unit, European Molecular Biology Laboratory; Heidelberg, Germany

†These authors contributed equally

\* Corresponding authors Emails: [ly2@sanger.ac.uk](mailto:ly2@sanger.ac.uk) (ORCID: <https://orcid.org/0000-0003-4519-7794>), [moritz.gerstung@ebi.ac.uk](mailto:moritz.gerstung@ebi.ac.uk), [mats.nilsson@scilifelab.se](mailto:mats.nilsson@scilifelab.se)

## Abstract

Subclonality is a universal feature of cancers yet how clones grow, are spatially organised, differ phenotypically or influence clinical outcome is unclear. To address this, we developed base specific in situ sequencing (BaSISS). In fixed tissues, transcripts harbouring clone-defining mutations are detected, converted into quantitative clone maps and characterised through multi-layered data integration. Applied to 8 samples from key stages of breast cancer progression BaSISS localised 1.42 million genotype informative transcripts across 4.9cm<sup>2</sup> of tissue. Microscopic clonal topographies are shaped by resident tissue architectures. Distinct transcriptional, histological and immunological features distinguish coexistent genetic clones. Spatial lineage tracing temporally orders clone features associated with the emergence of aggressive clinical traits. These results highlight the pivotal role of spatial genomics in deciphering the mechanisms underlying cancer progression.

## Introduction

Cancer growth is the result of mutation and selection of ever more proliferative clones analogous to Darwinian evolutionary theory(1–3). A consequence of this relentless process is that cancers are patchworks of genetically related but distinct groups of cells termed subclones(4, 5). While the somatic evolution model is well established due to the almost omnipresent existence of cancer subclones in bulk or multi-regional sequencing data(4–8), relatively little is currently known about the nature or causes of spatial patterns of cancer growth, phenotypic characteristics of distinct subclonal lineages or their interactions with the microenvironment. Still this information appears key because adverse cancer outcomes – growth, progression and recurrence – are properties of genetically distinct subclones (7, 9–11).

While a range of spatial molecular profiling strategies based on spatial RNA barcoding(12, 13), or fluorescence microscopy of single RNA molecules using different types of molecular probes exist(14–17), they do not perform the critical function of isolating genetic subclones in tissue context because gene expression profiles are highly plastic.

Evolutionary cancer genomics has demonstrated that lineage tracing using somatic mutations is a powerful and highly specific tool for tracing the subclonal origins of aggressive disease in earlier lesions(7, 18, 19). Histology driven sampling, such as laser capture microdissection(20) combined with low input nucleic acid library sequencing or even single cell sequencing goes some way towards resolving subclone spatial structure(21, 22). However, even the most exhaustive sampling strategy will struggle to provide an unbiased representation of the cancer landscape particularly across larger areas. Recently it has been demonstrated that molecular probes can be targeted to detect RNA in a sequence specific fashion, enabling detection of individual mutant transcripts *in situ*(15, 23). Still, these technologies have not been able to comprehensively map multiple clones, define spatial cancer evolution or subclone specific phenotypes.

To overcome these limitations, we have developed a Base Specific In Situ Sequencing (BaSISS) methodology that extends the In Situ Sequencing (ISS) protocol by incorporating multiplexed detection of clone specific mutations in fixed tissue specimens (24, 25). A dedicated Bayesian model then allows the interpretation of multi-layered spatial data in genetic clone-specific context. By applying the method to eight tissues from two multifocal breast cancers we generate the first ever large scale quantitative maps of cancer clones and three key messages emerge. 1) Patterns of spatial genetic heterogeneity are profoundly influenced by resident tissue structures; 2) Coexistent genetic clones can have distinct transcriptional, histological and immunological characteristics; 3) In preinvasive, invasive and locally metastatic breast tumours, the emergence of aggressive disease features can be temporally ordered and localised in genetic and histological contexts providing insights into the biology underlying cancer progression.

## Results

### **BaSISS detects bespoke panels of cancer-specific mutations in fixed tissues**

Cancer evolution produces multiple genetically related yet distinct clones each characterised by a unique combination of somatic mutations, known as the genotype, that

are related by the underlying phylogenetic tree (26). The spatial patterns created by coexisting cancer clones have not previously been directly observed. We established a pipeline to allow spatial mapping of cancer clones detected through standard bulk whole genome sequencing (WGS) and mutation clustering approaches (**Fig. 1A, Supp. Fig. S1, Supp. Methods**)(7, 9). Representative mutations from each cluster/branch of the phylogenetic tree were selected for spatial detection using BaSISS. The approach uses multiplexed highly sequence-specific padlock oligonucleotide probes with target recognition arms and 4-5 nucleotide reader barcodes to detect both mutant and wild-type alleles of each target (**Table S1**). Padlock probes bind to complementary DNA (cDNA), undergo rolling circle amplification and are read using sequencing by ligation with fluorophore-labelled interrogation probes (15, 27–29).

To assess whether spatial mutation signals provide a meaningful representation of the bulk genomic data, we first applied it to three different regions of a breast tumour obtained from the same mastectomy (whole breast) specimen (case PD9694) with confirmed inter- and intra-sample genetic heterogeneity defined by WGS (**Fig. 1B-E; Table S2-3**)(7). We designed 51 BaSISS probes to target representative genetic variants from the phylogenetic tree trunk and branches (tree annotations; **Fig. 1C, Table S1**). BaSISS was applied to 10 $\mu$ m thick, fresh frozen tissue sections up to 1.75cm in diameter derived from the same tissue blocks used for WGS. On average, 96% of BaSISS reads were on target and the median number of signals reporting each target was 3,719 (combined mutant and wild-type) across the 3 samples (**Table S4, Supp. Methods**). Plotting BaSISS mutation signals coloured according to the relevant branch, reveals inter-sample differences and these are largely consistent with bulk genomic data derived variant allele frequencies (VAF) ( $R=0.48-0.61$ , Pearson's) (**Fig. 1D, Fig. S2A-B**). Replicating the BaSISS experiment on serial tissue samples generates highly concordant sample-wise VAFs ( $R=0.76-0.93$ , Pearson's) and similar spatial signal distribution patterns (**Fig. S2C-D**). Spatial mutation and wild-type signal segregation patterns allowed the most likely phylogenetic tree, as represented in Figure 1C, to be selected from two solutions

previously considered equally likely based solely on WGS data (**Fig. S2D-E, Supp. Methods**).

### **Spatial genomics generates large scale cancer clone maps**

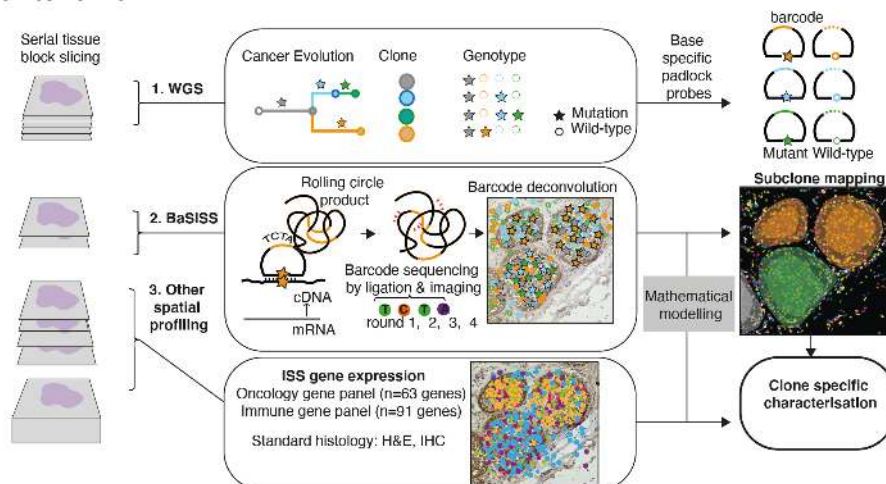
To generate continuous spatial subclone maps that quantify the local composition of cancer subclones and normal cells we developed a statistical algorithm that exploits BaSISS signals as well as local cell counts (derived from the DAPI channel during the fluorescence microscopy of BaSISS) using two dimensional Gaussian processes (**Fig. S1; Supp. Methods**). The variational Bayesian model also accounts for unspecific or wrongly decoded BaSISS signals and variable probe efficiency and is augmented by VAFs in bulk genomic sequencing data from serial tissue sections. The resulting maps of cancer (and normal) clones were reconstructed across a scale of up to 132mm<sup>2</sup> and with resolution of approximately 109 $\mu$ m (**Table S2**).

We applied the model to BaSISS data generated from eight tissue samples from 2 mastectomy specimens that together constitute 4.9cm<sup>2</sup> of breast tissue (**Fig. 1B, Table S2**). The cases were selected to chart the landmark stages of breast cancer progression. The first case, PD9694, consists of two discrete oestrogen receptor (ER) positive invasive primary breast cancers with ductal carcinoma in situ (DCIS). DCIS is a non-obligate precursor of invasive breast cancer consisting of neoplastic cells restricted to the duct lumen. Breach through the basement membrane into the breast stroma marks progression to invasive cancer. The second case, PD14780, includes two ER negative invasive breast cancers and a draining axillary lymph node that contains metastatic cells.

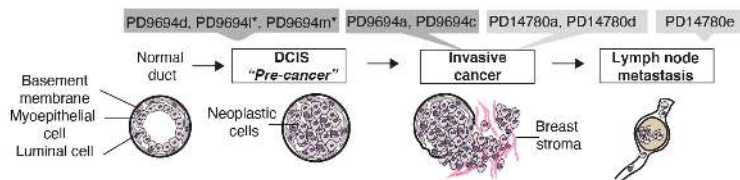
Mapping the dominant clone across each tissue reveals striking genetic variation within and between samples from the same case (**Fig. 2-4**). Comparison with histological appearances confirms that the model correctly differentiates between neoplastic regions (coloured fields) and normal tissues (white DAPI nuclei revealed; **Fig. 1E**, frequency plots reach zero; **Fig. 2A**). Applying the model to the validation BaSISS data generates highly consistent results (**Fig. S3A**).



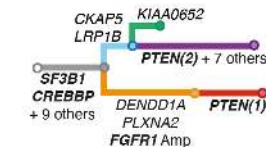
### A Spatial genomics workflow



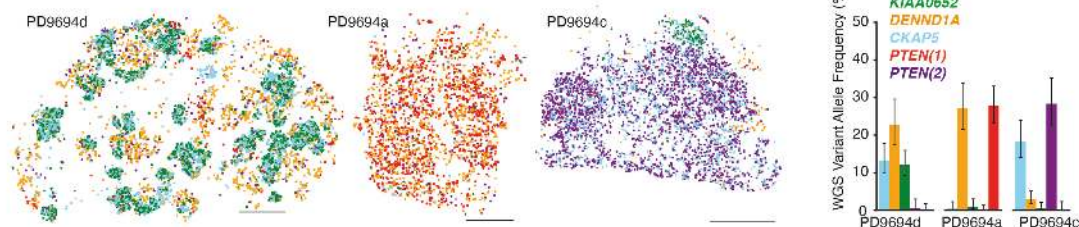
### B Key stages of breast cancer



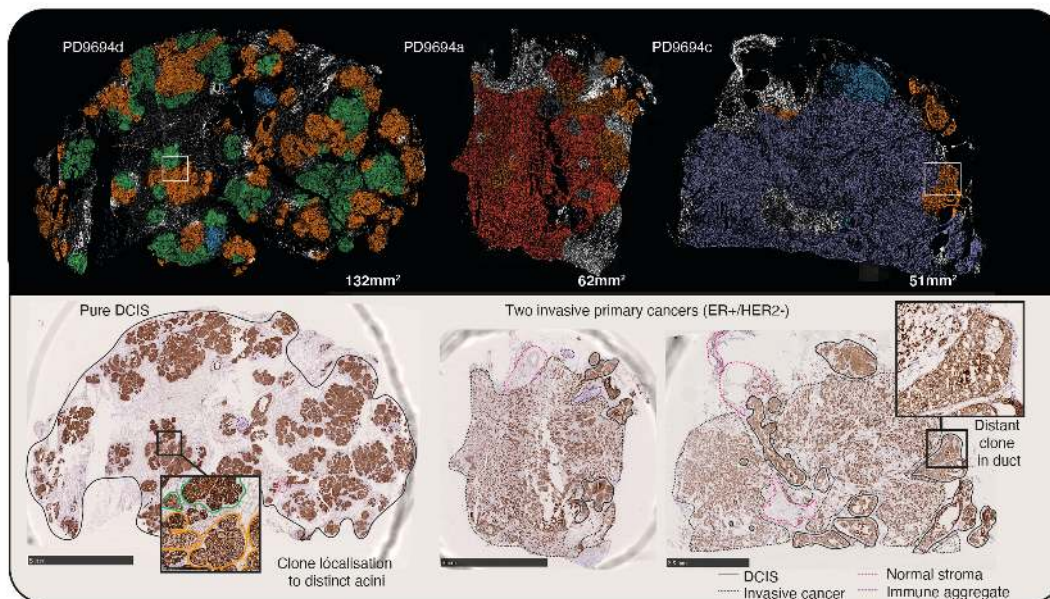
### C Case PD9694: Padlock probe targets by branch



### D BaSISS mutation signals



### E Cancer clone maps



**Fig. 1. Spatial Genomics generates maps of cancer clones** (A) The Spatial Genomics workflow detects whole genome sequencing (WGS) defined subclonal mutations within serial tissue sections using base specific in-situ sequencing (BaSISS). For mathematical aspects see Fig. S1. (B) To develop the spatial genomics approach we selected two cases where multiple samples (PD\*, shaded boxes) chart the distinct stages of primary breast cancer progression. Asterisk indicates no WGS data. (C) For case PD9694 most likely phylogenetic tree derived from WGS (samples PD9694a,c,d) and BaSISS data are presented with padlock probe targets named on each branch (driver mutations in bold). The barplot reveals the bulk genomic data (WGS plus targeted capture) derived variant allele frequency (VAF) of selected branch mutations in genes denoted in correspondingly coloured text (error bars report CIs = 5 - 95%). (D) BaSISS mutation signal plots – each dot represents the location of a mutation specific barcode, relating to the mutations reported in barplot (C). (E) Maps of the most prevalent clone projected on the DAPI image (reported if cancer cell fraction > 25% and inferred local cell density > 300 cells/mm<sup>2</sup>). Annotated pan-cytokeratin (pan-CK) IHC images of sequenced sections (bottom row, epithelial cells brown). Focus images show DCIS structure with coloured lines depicting approximate clone borders. DCIS = Ductal carcinoma in situ, ISS = gene expression in situ sequencing; IHC = immunohistochemistry, ER = oestrogen receptor, HER2 = Human epidermal growth factor receptor-2, cDNA = complementary DNA; mRNA = messenger RNA; H&E = haematoxylin and eosin.

### **Resident tissue structure influences subclonal growth patterns**

Examining the samples from the three stages of cancer progression together, it is evident that the resident tissue structure plays a fundamental role in defining the observed patterns of intratumoral heterogeneity. One consequence is the intimate juxtaposition of genetically distant clones. For example, the lymph node sample (PD14780e) that is examined in detail in relation to Figure 4, contains multiple clones but the sinus spaces are monopolised by a single clone. Similarly, the purple, invasive primary cancer PD9694c, is studded by evolutionarily distant orange clones but these are intraductal populations of neoplastic cells that are physically separated from the main cancer mass by the duct membrane (inset box, PD9694c; **Fig. 1E**). Similarly, in the three ‘pure’ DCIS samples, the microscopic structure of ducts and lobules underlies the striking, seemingly random mosaic of green and orange clones that are predicted to have diverged decades before (inset box, PD9694d; **Fig. 1E, Fig. 2A-B**).



The quantitative nature of these spatial genomics data allows us to further investigate the growth patterns in relation to tissue structure. In the pure DCIS samples (PD9694d,l,m), individual acinar and ductal ‘spaces’, defined by a myoepithelial cell layer and/or intervening stroma, are typically occupied by a complete clonal sweep (frequency plots; **Fig. 2A, Fig. S3B**). We deduce that this appearance is not simply a consequence of colonisation of distinct arms of the ductal system because we also observe distinct clonal fronts within several lobules that are corroborated by BaSISS mutation data, histological features and spatial transcriptomic data (**Fig. 2C-E**). These appearances might arise due to mutual tolerance or equal fitness, but in this case, as discussed below, the ability to spatially characterise co-existent clones provides evidence that we are more likely to be observing an incomplete clonal sweep by a fitter clone.

### **Spatial genomics reveals characteristic differences between DCIS clones with different fates**

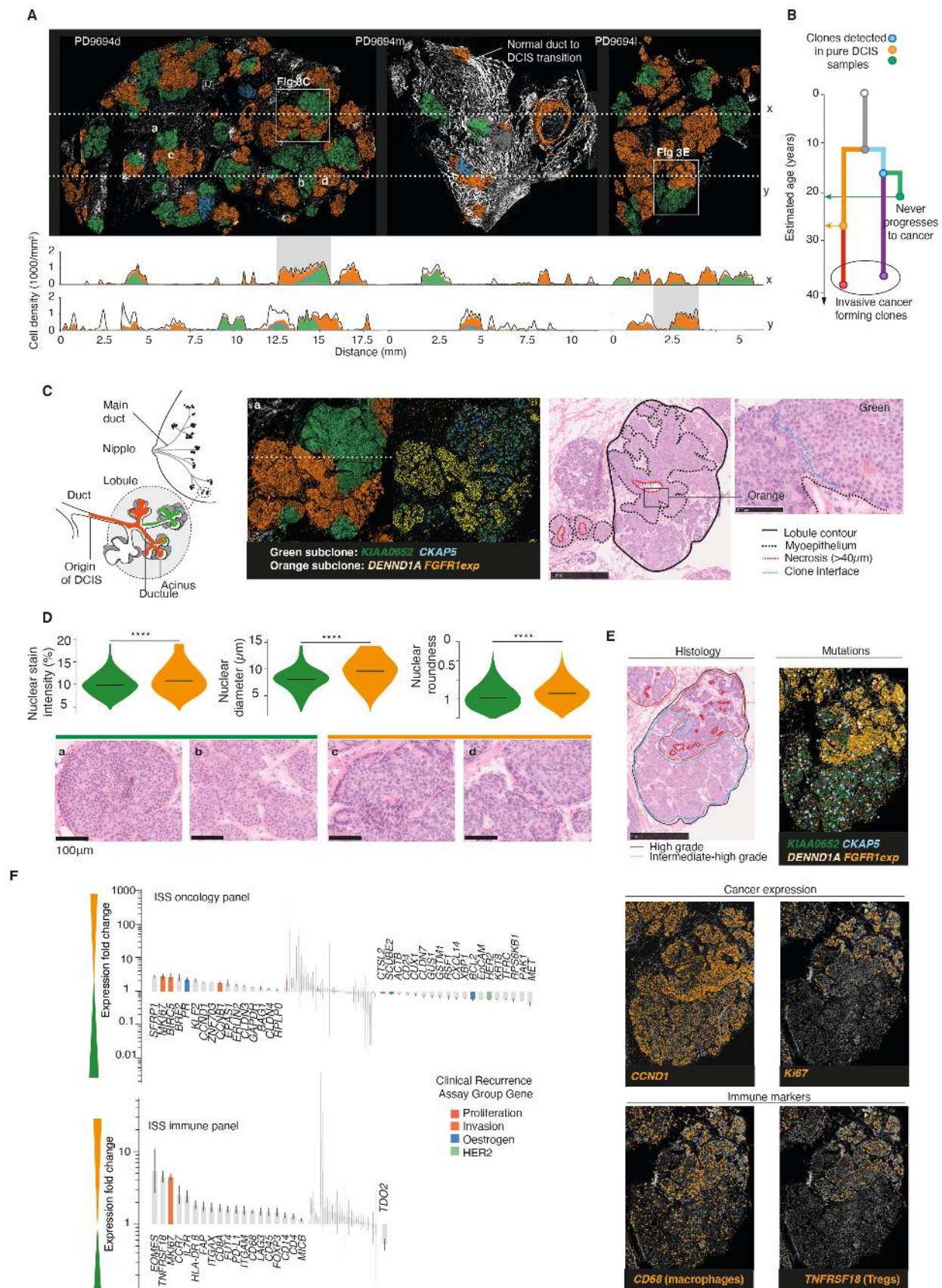
Why some, but not all DCIS lesions progress to a potentially lethal invasive cancer is poorly understood. This question is usually addressed by comparing cohorts of DCIS samples with different clinical outcomes but this approach is inherently difficult due to the heterogeneity of breast cancer as a disease(30, 31). Here, we demonstrate that spatial genomics provides a novel approach to the problem, by allowing the comparison of related DCIS clones that share many genetic and all host features yet manifest different clinical outcomes. In case PD9694, an invasive cancer arose from the orange clone but the green clone, despite predating orange by several years, never progressed (**Fig. 2B**).

Consistent with a more aggressive phenotype the orange clone cells have larger, more pleomorphic and more intensely stained nuclei as confirmed by digital pathology (**Fig. 2C-D**). Clone-specific histological features are remarkably stable, being recapitulated within virtually every distinct space and are also appreciated at the single nucleus level where clones meet, adding weight to BaSISS model inferences (**Fig. 2C-D, Fig. S3C**).

To derive clone-specific expression patterns, spatial transcriptomic data were generated on serial tissues sections using two bespoke ISS targeted panels: An immune marker panel (n=63 genes) and an oncology panel (n=91 genes) composed of a variety of breast epithelial markers and cancer pathway related genes including those derived from a clinical breast cancer recurrence assay (higher risk genes = proliferation, invasion, *HER2* group; lower risk genes = oestrogen group)(32, 33) (**Fig. 2F, Table S1, Table S5**). The per nucleus signal density was determined in regions dominated by either the orange or green subclone and significant differences determined (probability of positive log-ratio (PPLR) after Bonferroni correction  $< 0.01$ ) (**Table S6, Supp. Methods**). Relative to green, the ill-fated orange clone expresses higher levels of cell cycle regulatory oncogenes *CCND1*, *CCNB1*, the cell survival factor *Survivin* and the proliferative marker *Ki67* (1.8-4.4 fold) (**Fig. 2E-F, Fig. S3D**). Proliferative markers have demonstrated prognostic and predictive value in various clinical studies of DCIS progression risk (33–35).

Given the barrier between DCIS and the surrounding stroma, we were surprised to also observe that the majority of immune panel signals are enriched within the bounds of the orange clone fields. The most significantly enriched immune marker genes include those associated with T-regulatory cells (*TNFRSF18*, *ITGAM*), macrophages (*CD68*) and immune checkpoint inhibition (*PD-L1*, *LAG3*) (1.5-4.4 fold). The fibroblast marker (*CD34*) and fibroblast activation protein, *FAP*, signals are also denser in these areas (**Fig. 2F, Fig. S3E**). These findings are consistent with various studies that propose that the stromal environment plays an active role in driving DCIS progression but provides new evidence that this mechanism is clone specific (36–40).

Raw ISS signals overlaid on the histological image provide visual confirmation of the highly clone-specific nature of some signals and furthermore, the within clone variation in spatial localisation (**Fig. 2E**). It is remarkable that many of the features that differentiate between precancerous clones with different evolutionary trajectories foreshadow, albeit at lower amplitude, many of the changes that distinguish invasive cancers from preinvasive cancers in general (**Fig. 3**) (36, 41, 42).



**Fig. 2. Histological, transcriptional and immunological differences between DCIS clones with different fates** (A) Maps of pure DCIS clones (as for Fig. 1E) and frequency plots of local, mean cancer (coloured areas) and non-cancer (white) composition, corresponding to horizontal dashed lines. (B) WGS derived phylogenetic tree reports relative timing of emergence (arrows) and evolutionary fates of DCIS clones in case PD9694. Timing estimates are derived from branch mutation burdens and patient age and assume steady mutation rates. (C) Pure DCIS clone growth patterns: Cartoon illustrates anatomical organisation of breast duct and lobule system and possible green/orange clone locations. Other images provide detailed views of segregated clone growth within a distended lobule in PD9694d (see A): clone fields (left) and select BaSISS mutation signals (middle) on DAPI image (coloured by clone: *FGFR1* is an expression probe so is more highly expressed but not exclusive to the orange subclone, other probes are mutation specific); annotated H&E stained serial section with focus image demonstrating different appearances across clone interface. (D) Histological appearances (H&E stained serial sections) of select areas corresponding to letters a-d in (A) reveal larger, more pleomorphic and intensely stained nuclei in orange clone. Violin plots report nuclear morphological features of green (11,365 nuclei) and orange (11,699 nuclei) regions (all differences significant,  $p < 0.0001$ , Mann-Whitney U). (E) BaSISS and ISS data projected on the DAPI image reveal orange clone specific characteristics in a selected area of PD9694(A). (F) Barplots report fold changes in ISS gene panel expression signals between orange (1,515 tiles) and green clone regions (2,583 tiles) (predominant clone defines the location). Annotated genes have significant expression differences (probability of positive log-ratio (PPLR) after Bonferroni correction  $< 0.01$ ). Genes are ordered by combined PPLR and direction of change (fine lines report CIs = 5 - 95%).

### Temporal ordering of genetic, transcriptional and immunological changes during the development of invasive cancer

How DCIS progresses to an invasive cancer is poorly understood. In the previous section we observed characteristic differences between competing clones in pure DCIS samples and here we extend this by tracing the changes accompanying the progression to invasive cancer – a potentially fatal condition (**Fig. 3A**). The two separate invasive cancers (PD9694a and PD9694c) arose from a common precursor clone along divergent lineages – orange to red and blue to purple – with parallel loss of *PTEN* through distinct damaging mutations (**Fig. 3A-B**). *PTEN* loss is the only invasion specific driver mutation identified within each lineage and was evidently under strong selection during progression as has been demonstrated in other breast cancers (43). In both lineages, the *PTEN* mutant clone



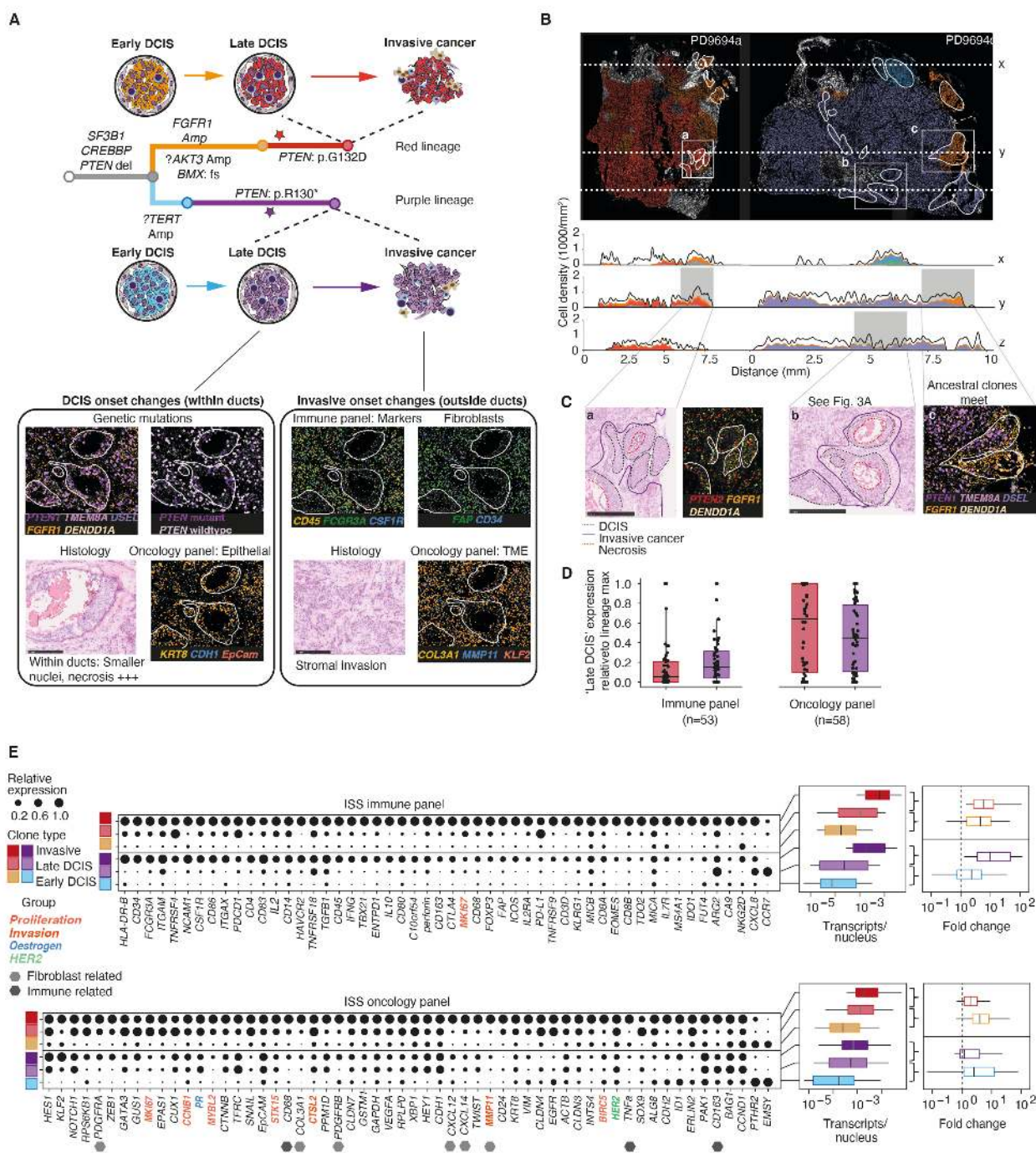
dominates the invasive cancer but unexpectedly, also accounts for sizable areas of DCIS consistent with an intraductal onset (**Fig. 3B-C**). The terminal purple lineage clone is differentiated by 8 private mutations and these are all present at similar levels in the DCIS and invasive compartments supporting previous reports of genetic similarity between the two disease states (21, 30) (**Fig. S4A**). Histological appearances are more consistent with this reflecting genetic evolution within the ducts rather than widespread recolonisation by the invasive cancer and is referred to herein as ‘late DCIS’ to differentiate it from the ancestral DCIS clone (early DCIS). By comparing the targeted ISS gene expression data in early DCIS, late DCIS (with *PTEN* mutations), and invasive cancer clones along each lineage we sought to temporally order when the non-genetic features associated with an invasive cancer emerge (**Fig. 3A**). Here, we investigate to which extent transcriptional changes are replicated by the two lineages that display convergent genetic evolution in the same host and same genetic background.

For each lineage, gene expression profiles were determined for the early DCIS, late DCIS and invasive cancer regions using targeted ISS. For 53 immune panel and 58 oncology panel genes, a significant difference in expression is evident between early DCIS and invasive cancer regions (PPLR after Bonferroni correction  $< 0.01$ ) (**Fig. 2D-E, Table S6**). For most genes the expression level is higher in the invasive cancer (**Fig. S4B**). The expression signal density in late DCIS is usually somewhere between early DCIS and invasive cancer, although differences exist between genes. Most immune panel genes behave in a similar fashion, exhibiting less than 25% of the signal enrichment detected in the invasive cancer (**Fig. 3D-E**). This is consistent with the picture whereby immune cells, by and large, remain outside of the ducts irrespective of genetic features of the clone inside (**Fig. 3A**, box b; **Fig. S4C**). In late DCIS, some immune cell signals appear clustered close to ‘microbreaches’ in the mesenchymal cell layer (arrowheads, box c; **Fig. S4C**). Interestingly, in one region, immune signals cluster specifically with a group of purple clone cells in a duct that is otherwise occupied by an evolutionarily distant orange clone – a picture that, in the absence of the precursor DCIS clone, might support a pattern of cancer reinvasion (far right box; **Fig. 3C, Fig. S4D**). A smaller group of immune marker

genes including T-regulatory cell marker (*TNFRSF18*) and the mainly cancer cell derived MHC class I chain related-proteins A and B (*MICA*, *MICB*) do not exhibit clustering but are diffusely scattered throughout the purple DCIS. This pattern is also apparent in the pure DCIS comparisons (orange versus green) in the previous section, it is therefore conceivable that these more diffuse immune changes might gradually emerge during waves of clonal progression (**Fig. 2E-F**).

The oncology panel genes are more heterogeneous in their expression behaviour and this undoubtedly reflects the more diverse composition of the panel. For around half of the genes, at least 50% of the increase in expression level seen in the invasive cancer is evident within the late DCIS (left box; **Fig. 3A**, **Fig. 3D-E**). A range of genes, including prognostic genes linked to both proliferation (*Ki67*) and invasion (*CTSL2*), have similar expression levels in late DCIS and invasive clones (**Fig. 3E**)(32). These gene signals are diffusely distributed throughout the purple clone regions without clustering indicating they are a general property of the DCIS cells. It therefore appears that at the DCIS stage, neoplastic cells can be armed with both the genetic and many of the transcriptional changes needed for life as an invasive cancer. It is perhaps remarkable that these cells form extensive expansions within intact ducts rather than invading immediately. While these data support the model of intraductal clonal progression they also indicate that the ultimate invasive transition could well be driven by the tumour microenvironment or rapid expansion of one clone could physically rupture the ductal basement membrane, spilling multiple clones into the breast stroma (21, 30).





**Fig. 3. Temporal ordering of genetic, transcriptional and immunological changes during the development of invasive cancer. (A)** Cartoon illustrates parallel lineages of progression from DCIS to invasive cancer. Potential driver mutations annotate relevant branches of the phylogenetic tree. DCIS onset changes (left box) include *PTEN* driver mutations and expression of genes typically associated with epithelial/cancer origin. Invasive onset changes (right box) include expression of genes associated with the

tumour microenvironment (TME). **(B)** Clone maps (as per Fig. 1E) and frequency plots of invasive cancers with DCIS regions marked in white: *PTEN* mutant clones (red/purple) occupy several mm<sup>2</sup> of histologically confirmed DCIS (A and **C**). **(D)** Boxplots report, for the genes upregulated between early DCIS and invasive cancer, the extent to which this is acquired within late DCIS. Significantly altered genes are defined as having PPLR after Bonferroni correction  $< 0.01$ . **(E)** A dot plot showing the relative expression of the same genes from (D) at the different progression stages in the 2 lineages. Dot area = (transcript/nucleus) divided by maximum value for each gene. Boxplots report transcripts/ nucleus (left) for genes shown and fold change between indicated comparisons (right). For all boxplots we report median, lower and upper quartiles (box) and 5-95 percentiles (whiskers). Expression per clone was defined on 136 - 3262 tiles, see Table S6 for details.

### **Tracing the emergence of aggressive cancer characteristics in a lymph node metastasis**

Lymph node metastasis predicts distant metastasis and death, but whether it plays an active role in facilitating cancer progression or simply reflects more aggressive primary cancer biology is unknown. To assess whether the lymph node might drive clinically meaningful evolutionary progression, we selected a case where the clinically targetable breast cancer oncogene *HER2* was found to be amplified in the lymph node but not primary tumours by WGS (9) (**Fig. 4A, Fig. S5A**). As in case 1, we designed BaSISS padlock probes to mutations from the branches of the WGS inferred tree (stars; **Fig. 4A**). Targets included a passenger mutation in *HER2* and a novel internal ‘fusion’ in *CACNB1* (a gene 5 prime to *HER2*). These mutations are predicted to have occurred prior to and during the breakage fusion bridge event that generated the *HER2* amplification event respectively, and were included to aid evolutionary timing (**Fig. 4A, Fig. S5A, Table S1**).

The BaSISS signal data exhibit spatial patterns that support the existence of at least 2 lymph node clones: A post-amplification clone with high *HER2* mutation density and *CACNB1* signals and at least one pre-amplification clone with low level *HER2* mutation signals but similar trunk mutation density (BaSISS plots; **Fig. 4B-C**). This supports WGS subclonal copy number data (**Fig. S5A**). We therefore provided the model with a genotype

matrix for the 4 clones indicated in **Figure 4A (Table S1)**, generating the clone maps of primary cancers and the lymph node in **Figures 4D-E**.

A subclonal population (purple) is detected in one of the primary cancers and this exhibits histological and expression features of more aggressive disease compared to the more ancestral (blue) clone (**Fig. S4B-D**). However, we did not find any evidence to support the emergence of the *HER2* amplified (orange) clone or the *HER2* mutation bearing, pre-amplification (green) clone within the primary cancers here or on a serial tissue section. Nonetheless, the existence of multiple clones in the lymph node confirms that *HER2* amplification was not necessary to establish metastasis in this cancer. Although we cannot exclude the possibility that *HER2* amplification occurred in an unsampled region of the primary tumour, the presence of the green *HER2* mutant, pre-amplification clone in the lymph node is consistent with emergence at this site (**Fig. 4B-C**).

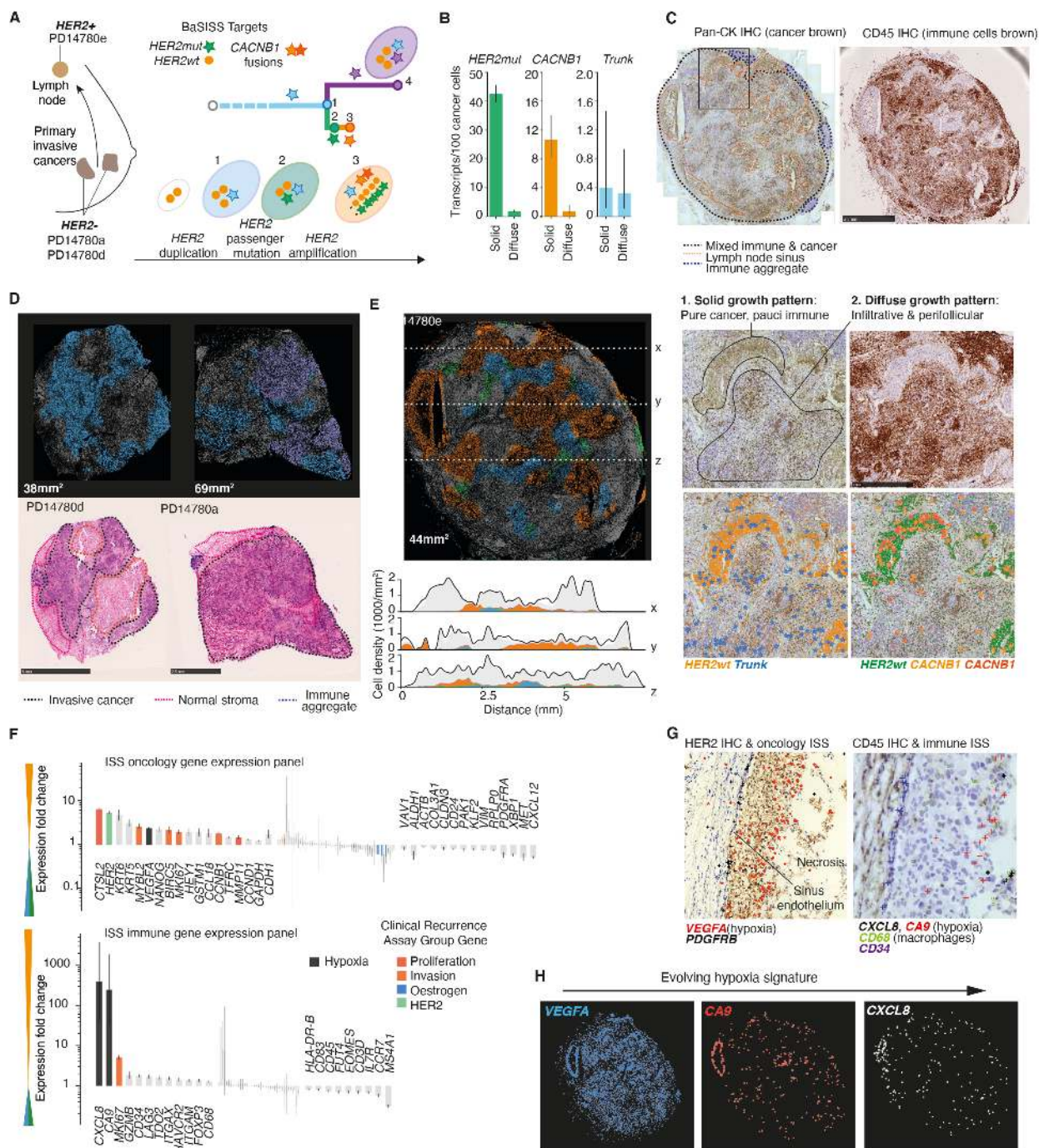
The lymph node contains more than one pattern of cancer growth and these patterns are associated with different genetic subclones. The first, a solid growth pattern composed of pure cancer cells with no obvious stromal component and a paucity of immune cells is entirely formed by the orange clone (**Fig. 4C, Fig. 4E**). Orange deposits are surrounded by endothelial cell layers and neat lines of *CD34* and *PDGFRB* ISS signals that are expressed by lymph node sinuses and potentially microvessels (**Fig. 4G, Fig. S5E**). The remainder of the node is heavily infiltrated by cancer cells that exhibit a diffuse infiltrative pattern, intermingling with immune cells and frequently forming perifollicular aggregates (**Fig. 4C, Fig. S5F**). Most of these cells are assigned to one or other pre-*HER2*-amplification clone, although confidence of exact clone assignment is low due to low cancer purity in these areas. We therefore observe that the lymph node offers a distinct environment that permits segregated growth of genetically distinct subclones.

Consistent with a more aggressive emerging subclone, the orange subclone expresses higher levels of invasion and proliferation related genes, relative to the green and blue subclones (32) (**Fig. 4F; Supp. Methods**). Interestingly, three of the most highly enriched

genes – Vascular endothelial growth factor A (*VEGFA*) (2.4 fold), carbonic anhydrase IX (*CA9*)(236 fold) and C-X-C motif chemokine ligand 8 (*CXCL8*)(384 fold) – are implicated in hypoxia and angiogenesis and their ISS signals follow striking patterns (PPLR =0.001) (44-46) (**Fig. 4G-H**). The signals are highly specific to the orange subclone and their acquisition is cumulative, following metastatic deposit size. For each gene, density within a deposit is inversely related to the distance from the endothelium and presumably the inverse oxygen gradient (**Fig. 4G**). While *CA9* and *VEGFA* seem to be mainly cancer cell derived, *CXCL8* co-localises with CD45+ immune/lymphoid cells (by IHC) and *CD68/CD163* signals (by ISS) consistent with a macrophage origin in this scenario (**Fig. S4G**). Both *HER2* amplification and hypoxic signatures are associated with adverse clinical outcomes in breast cancer (47). This case provides another powerful example of how spatial genomics can uncover the interplay of genetics and tumour microenvironment in contributing towards the subclonal diversification that underpins the emergence of clinically aggressive cancers.

While the eight samples interrogated with BaSISS revealed a wealth of information related to their individual patterns of subclonal growth, subclone-specific phenotypes and successive transcriptomic changes acquired during progression, some overarching patterns also emerge. Evaluating the clinical recurrence assay gene groups based on the prognostic ISS marker panel, we observe that proliferative and invasive group gene scores steadily increase along seven predicted clonal succession episodes spanning three cancer stages from DCIS to lymph node metastasis (**Fig. S6**). This would be consistent with the notion that major clonal succession events in a cancer reflect the emergence of increasingly proliferative clones (48). Larger spatial genomics studies may guide the development of precise biomarkers and molecular staging systems by exploiting the sequential nature of subclone specific changes.





**Fig. 4. Tracing aggressive disease in a lymph node metastasis. (A)** Case PD14780 samples, phylogenetic tree and genotypes (ovals represent a cell) inferred from WGS data and BaSIS signals. **(B)** Barplot of BaSIS signals extracted from 939 and 894 tiles in solid and diffuse growth pattern regions provide evidence of at least two genotypes: One with amplified *HER2* mutation signals and *CACNB1* fusion signals, the other with low level/ absent *HER2* mutation signals but similar trunk mutation signal density (CIs = 5-95%). ‘Trunk’ comprises 3 different mutations. **(C)** Sample PD14780e, BaSIS/ ISS sequenced

lymph node sections stained by IHC with pan-cytokeratin (pan-CK, stains cancer cells brown) and CD45 (stains immune cells brown). Whole section (top row), focus area (middle row) and focus area (pan-CK stained) with overlaid BaSISS signals. **(D-E)** BaSISS subclone fields (top row) (DAPI projected, colour reflects the most prevalent subclone, reported if cancer cell fraction > 15% (D) and > 25% (E)) and (bottom row) annotated H&E stained serial sections (D) or subclone frequency plots (E). **(F)** Barplots of subclone specific gene expression fold changes in PD14780e (orange versus blue/green aggregated signals) for immune and oncology panels ordered by PPLR and direction of change. Genes with significant expression difference (PPLR after Bonferroni correction < 0.01) are annotated. **(G)** Spatial patterns of *PDGFRB*, *CD34* and hypoxia related ISS signals in PD14780e overlaid on HER2 (left) and CD45 IHC stained sections(right). **(H)** Spatial patterns of 3 hypoxia related genes show spatial patterns and are cumulative.

## Discussion

Here we present BaSISS, a highly multiplexed fluorescence microscopy based protocol to map and phenotypically characterise cancer clones. The BaSISS technology supports a tailored approach according to each cancer's unique complement of mutations. In this proof of principle study we map relatively broad subclone populations identified through multi-region WGS but the spatial genomics approach could equally be applied to more detailed phylogenies. A particular advantage of the technology is that it is capable of interrogating very large tissue sections and that it is comparably cheap, unlike solely relying on sequencing based methods (49). In theory, the approach holds the potential to create three dimensional genomic tomographs by aligning consecutive tissue sections. A limitation of the approach is relatively low sensitivity, which currently precludes single cell genotyping. For standard ISS, additional sensitivity can be achieved by tiling transcripts with more probes; unfortunately, this is not feasible for point mutations at a defined genomic location. A switch to hybridisation based sequencing and direct RNA binding probes, which eliminate the requirement for reverse transcription are currently limited to gene expression, but with further development should also improve base specific detection several fold (50, 51).

BaSISS' ability to spatially locate and molecularly characterise different cancer subclones adds essential features to the spatial genomics tool kit. 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 application of spatial genomics approaches will uncover how cancers grow in different tissues and allow us to track, trace and characterise the ill-fated clones that are responsible for adverse clinical outcomes.

## **Acknowledgements**

### **Funding**

LRY is funded by a Wellcome Trust Clinical Research Career Development Fellowship ref: 214584/Z/18/Z. This work was supported by a pump-priming award from the Cancer Research UK Cambridge Centre Early Detection Programme [CRUK grant ref: A25117]. MN is funded by the Swedish Research Council (project grant 2019-01238), Cancerfonden (project grant CAN 2018/604), and the strategic research area U-CAN.

### **Author contributions**

JS, PJC, MN and LRY designed the study. AL, MG and LRY analysed and interpreted data and drafted the article and figures. AL and MG developed the core mathematical models. AS and VK contributed to mathematical modeling. AS contributed to image segmentation and processing. JS, MN and CS acquired ISS data and contributed to data interpretation and manuscript preparation. AR provided samples. ALR, LM, SP contributed histopathological expertise. YSJ contributed RNAseq expertise. SD contributed to WGS sub clonality analysis. JSP, VV, TL, OAB, MGataric contributed to development of bespoke ISS analysis pipelines. All authors reviewed and commented on the manuscript.

### **Competing interests**

CS is co-owner of HistoOne AB, Sweden, and has research contracts with Prelude Dx, CA, US. MN is an advisor to 10X Genomics. JS is now (but was not at the time of contribution to this manuscript) an employee of Spatial Transcriptomics, Part of 10x Genomics, Inc, Södra Fiskartorpsvägen 15C, 114 33 Stockholm, Sweden. Other authors

declare no competing interests.

## Data and materials availability

Code that was used in the analysis could be found in <https://github.com/gerstung-lab/BaSISS>. Code used to segment nuclei in images is available in <https://github.com/yozhikoff/segmentation>. Data is available upon request.

## List of Supplementary Materials

Materials and Methods

Fig. S1-S6

Table S1-S6

Methods references (7, 9, 15, 28, 32, 33, 52–65)

## References

1. J. Cairns, Mutation selection and the natural history of cancer. *Nature*. **255**, 197–200 (1975).
2. P. C. Nowell, The clonal evolution of tumor cell populations. *Science*. **194**, 23–28 (1976).
3. L. R. Yates, P. J. Campbell, Evolution of the cancer genome. *Nat. Rev. Genet.* **13**, 795–806 (2012).
4. S. C. Dentro, I. Leshchiner, K. Haase, M. Tarabichi, J. Wintersinger, A. G. Deshwar, K. Yu, Y. Rubanova, G. Macintyre, J. Demeulemeester, I. Vázquez-García, K. Kleinheinz, D. G. Livitz, S. Malikic, N. Donmez, S. Sengupta, P. Anur, C. Jolly, M. Cmero, D. Rosebrock, S. Schumacher, Y. Fan, M. Fittall, R. M. Drews, X. Yao, J. Lee, M. Schlesner, H. Zhu, D. J. Adams, G. Getz, P. C. Boutros, M. Imielinski, R. Beroukhim, S. Cenk Sahinalp, Y. Ji, M. Peifer, I. Martincorena, F. Markowitz, V. Mustonen, K. Yuan, M. Gerstung, P. T. Spellman, W. Wang, Q. D. Morris, D. C. Wedge, P. Van Loo, on behalf of the PCAWG Evolution and Heterogeneity Working Group the PCAWG consortium, the PCAWG consortium, Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. *Cold Spring Harbor Laboratory* (2020), p. 312041.
5. N. Andor, T. A. Graham, M. Jansen, L. C. Xia, C. A. Aktipis, C. Petritsch, H. P. Ji, C. C. Maley, Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat. Med.* **22**, 105–113 (2016).

6. M. Gerlinger, A. J. Rowan, S. Horswell, J. Larkin, D. Endesfelder, E. Gronroos, P. Martinez, N. Matthews, A. Stewart, P. Tarpey, I. Varela, B. Phillimore, S. Begum, N. Q. McDonald, A. Butler, D. Jones, K. Raine, C. Latimer, C. R. Santos, M. Nohadani, A. C. Eklund, B. Spencer-Dene, G. Clark, L. Pickering, G. Stamp, M. Gore, Z. Szallasi, J. Downward, P. A. Futreal, C. Swanton, Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
7. L. R. Yates, M. Gerstung, S. Knappskog, C. Desmedt, G. Gudem, P. Van Loo, T. Aas, L. B. Alexandrov, D. Larsimont, H. Davies, Y. Li, Y. S. Ju, M. Ramakrishna, H. K. Haugland, P. K. Lilleng, S. Nik-Zainal, S. McLaren, A. Butler, S. Martin, D. Glodzik, A. Menzies, K. Raine, J. Hinton, D. Jones, L. J. Mudie, B. Jiang, D. Vincent, A. Greene-Colozzi, P.-Y. Adnet, A. Fatima, M. Maetens, M. Ignatiadis, M. R. Stratton, C. Sotiriou, A. L. Richardson, P. E. Lønning, D. C. Wedge, P. J. Campbell, Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat. Med.* **21**, 751–759 (2015).
8. R. Sun, Z. Hu, A. Sottoriva, T. A. Graham, A. Harpak, Z. Ma, J. M. Fischer, D. Shibata, C. Curtis, Between-region genetic divergence reflects the mode and tempo of tumor evolution. *Nat. Genet.* **49**, 1015–1024 (2017).
9. L. R. Yates, S. Knappskog, D. Wedge, J. H. R. Farmery, S. Gonzalez, I. Martincorena, L. B. Alexandrov, P. Van Loo, H. K. Haugland, P. K. Lilleng, G. Gudem, M. Gerstung, E. Pappaemmanuil, P. Gazinska, S. G. Bhosle, D. Jones, K. Raine, L. Mudie, C. Latimer, E. Sawyer, C. Desmedt, C. Sotiriou, M. R. Stratton, A. M. Sieuwerts, A. G. Lynch, J. W. Martens, A. L. Richardson, A. Tutt, P. E. Lønning, P. J. Campbell, Genomic Evolution of Breast Cancer Metastasis and Relapse. *Cancer Cell.* **32**, 169–184.e7 (2017).
10. C. C. Maley, P. C. Galipeau, J. C. Finley, V. J. Wongsurawat, X. Li, C. A. Sanchez, T. G. Paulson, P. L. Blount, R.-A. Risques, P. S. Rabinovitch, B. J. Reid, Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nat. Genet.* **38**, 468–473 (2006).
11. M. Jamal-Hanjani, G. A. Wilson, N. McGranahan, N. J. Birkbak, T. B. K. Watkins, S. Veeriah, S. Shafi, D. H. Johnson, R. Mitter, R. Rosenthal, M. Salm, S. Horswell, M. Escudero, N. Matthews, A. Rowan, T. Chambers, D. A. Moore, S. Turajlic, H. Xu, S.-M. Lee, M. D. Forster, T. Ahmad, C. T. Hiley, C. Abbosh, M. Falzon, E. Borg, T. Marafioti, D. Lawrence, M. Hayward, S. Kolvekar, N. Panagiotopoulos, S. M. Janes, R. Thakrar, A. Ahmed, F. Blackhall, Y. Summers, R. Shah, L. Joseph, A. M. Quinn, P. A. Crosbie, B. Naidu, G. Middleton, G. Langman, S. Trotter, M. Nicolson, H. Remmen, K. Kerr, M. Chetty, L. Gomersall, D. A. Fennell, A. Nakas, S. Rathinam, G. Anand, S. Khan, P. Russell, V. Ezhil, B. Ismail, M. Irvin-Sellers, V. Prakash, J. F. Lester, M. Kornaszewska, R. Attanoos, H. Adams, H. Davies, S. Dentro, P. Taniere, B. O’Sullivan, H. L. Lowe, J. A. Hartley, N. Iles, H. Bell, Y. Ngai, J. A. Shaw, J. Herrero, Z. Szallasi, R. F. Schwarz, A. Stewart, S. A. Quezada, J. Le Quesne, P. Van Loo, C. Dive, A. Hackshaw, C. Swanton, Tracking the Evolution of Non-Small-

- Cell Lung Cancer. *N. Engl. J. Med.* **376**, 2109–2121 (2017).
12. P. L. Ståhl, F. Salmén, S. Vickovic, A. Lundmark, J. F. Navarro, J. Magnusson, S. Giacomello, M. Asp, J. O. Westholm, M. Huss, A. Mollbrink, S. Linnarsson, S. Codeluppi, Å. Borg, F. Pontén, P. I. Costea, P. Sahlén, J. Mulder, O. Bergmann, J. Lundeberg, J. Frisén, Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*. **353**, 78–82 (2016).
  13. S. G. Rodrigues, R. R. Stickels, A. Goeva, C. A. Martin, E. Murray, C. R. Vanderburg, J. Welch, L. M. Chen, F. Chen, E. Z. Macosko, Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science*. **363**, 1463–1467 (2019).
  14. F. Wang, J. Flanagan, N. Su, L.-C. Wang, S. Bui, A. Nielson, X. Wu, H.-T. Vo, X.-J. Ma, Y. Luo, RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J. Mol. Diagn.* **14**, 22–29 (2012).
  15. R. Ke, M. Mignardi, A. Pacureanu, J. Svedlund, J. Botling, C. Wählby, M. Nilsson, In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods*. **10**, 857–860 (2013).
  16. S. Codeluppi, L. E. Borm, A. Zeisel, G. La Manno, J. A. van Lunteren, C. I. Svensson, S. Linnarsson, Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat. Methods*. **15**, 932–935 (2018).
  17. C.-H. L. Eng, M. Lawson, Q. Zhu, R. Dries, N. Koulena, Y. Takei, J. Yun, C. Cronin, C. Karp, G.-C. Yuan, L. Cai, Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature*. **568**, 235–239 (2019).
  18. S. Jones, W.-D. Chen, G. Parmigiani, F. Diehl, N. Beerenwinkel, T. Antal, A. Traulsen, M. A. Nowak, C. Siegel, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, J. Willis, S. D. Markowitz, Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 4283–4288 (2008).
  19. S. P. Shah, R. D. Morin, J. Khattra, L. Prentice, T. Pugh, A. Burleigh, A. Delaney, K. Gelmon, R. Guliany, J. Senz, C. Steidl, R. A. Holt, S. Jones, M. Sun, G. Leung, R. Moore, T. Severson, G. A. Taylor, A. E. Teschendorff, K. Tse, G. Turashvili, R. Varhol, R. L. Warren, P. Watson, Y. Zhao, C. Caldas, D. Huntsman, M. Hirst, M. A. Marra, S. Aparicio, Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature*. **461**, 809–813 (2009).
  20. C. Y. Shen, J. C. Yu, Y. L. Lo, C. H. Kuo, C. T. Yue, Y. S. Jou, C. S. Huang, J. C. Lung, C. W. Wu, Genome-wide search for loss of heterozygosity using laser capture microdissected tissue of breast carcinoma: an implication for mutator phenotype and breast cancer pathogenesis. *Cancer Res.* **60**, 3884–3892 (2000).
  21. A. K. Casasent, A. Schalck, R. Gao, E. Sei, A. Long, W. Pangburn, T. Casasent, F. Meric-Bernstam, M. E. Edgerton, N. E. Navin, Multiclonal Invasion in Breast Tumors

- Identified by Topographic Single Cell Sequencing. *Cell*. **172**, 205–217.e12 (2018).
22. E. Laks, A. McPherson, H. Zahn, D. Lai, A. Steif, J. Brimhall, J. Biele, B. Wang, T. Masud, J. Ting, D. Grewal, C. Nielsen, S. Leung, V. Bojilova, M. Smith, O. Golovko, S. Poon, P. Eirew, F. Kabeer, T. Ruiz de Algara, S. R. Lee, M. J. Taghiyar, C. Huebner, J. Ngo, T. Chan, S. Vatrt-Watts, P. Walters, N. Abrar, S. Chan, M. Wiens, L. Martin, R. W. Scott, T. M. Underhill, E. Chavez, C. Steidl, D. Da Costa, Y. Ma, R. J. N. Coope, R. Corbett, S. Pleasance, R. Moore, A. J. Mungall, C. Mar, F. Cafferty, K. Gelmon, S. Chia, CRUK IMAXT Grand Challenge Team, M. A. Marra, C. Hansen, S. P. Shah, S. Aparicio, Clonal Decomposition and DNA Replication States Defined by Scaled Single-Cell Genome Sequencing. *Cell*. **179**, 1207–1221.e22 (2019).
  23. A.-M. Baker, W. Huang, X.-M. M. Wang, M. Jansen, X.-J. Ma, J. Kim, C. M. Anderson, X. Wu, L. Pan, N. Su, Y. Luo, E. Domingo, T. Heide, A. Sottoriva, A. Lewis, A. D. Beggs, N. A. Wright, M. Rodriguez-Justo, E. Park, I. Tomlinson, T. A. Graham, Robust RNA-based in situ mutation detection delineates colorectal cancer subclonal evolution. *Nat. Commun.* **8**, 1998 (2017).
  24. R. Ke, M. Mignardi, A. Pacureanu, J. Svedlund, J. Botling, C. Wählby, M. Nilsson, In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods*. **10**, 857–860 (2013).
  25. C. Larsson, I. Grundberg, O. Söderberg, M. Nilsson, In situ detection and genotyping of individual mRNA molecules. *Nat. Methods*. **7**, 395–397 (2010).
  26. S. González, N. Volkova, P. Beer, M. Gerstung, Immuno-oncology from the perspective of somatic evolution. *Semin. Cancer Biol.* **52**, 75–85 (2018).
  27. X. Qian, K. D. Harris, T. Hauling, D. Nicoloutsopoulos, A. B. Muñoz-Manchado, N. Skene, J. Hjerling-Leffler, M. Nilsson, Probabilistic cell typing enables fine mapping of closely related cell types in situ. *Nat. Methods* (2019), doi:10.1038/s41592-019-0631-4.
  28. J. Svedlund, C. Strell, X. Qian, K. J. C. Zilkens, N. P. Tobin, J. Bergh, A. M. Sieuwerts, M. Nilsson, Generation of in situ sequencing based OncoMaps to spatially resolve gene expression profiles of diagnostic and prognostic markers in breast cancer. *EBioMedicine*. **48**, 212–223 (2019).
  29. C. Yokota, D. Gyllborg, M. Nilsson, In situ sequencing for RNA analysis in tissue sections v2 (protocols.io.bb2giqbw), , doi:10.17504/protocols.io.bb2giqbw.
  30. F. Pareja, D. N. Brown, J. Y. Lee, A. Da Cruz Paula, P. Selenica, R. Bi, F. C. Geyer, A. Gazzo, E. M. da Silva, M. Vahdatinia, A. A. Stylianou, L. Ferrando, H. Y. Wen, J. B. Hicks, B. Weigelt, J. S. Reis-Filho, Whole-Exome Sequencing Analysis of the Progression from Non-Low-Grade Ductal Carcinoma In Situ to Invasive Ductal Carcinoma. *Clin. Cancer Res.* **26**, 3682–3693 (2020).

31. S. Nik-Zainal, H. Davies, J. Staaf, M. Ramakrishna, D. Glodzik, X. Zou, I. Martincorena, L. B. Alexandrov, S. Martin, D. C. Wedge, P. Van Loo, Y. S. Ju, M. Smid, A. B. Brinkman, S. Morganella, M. R. Aure, O. C. Lingjærde, A. Langerød, M. Ringnér, S.-M. Ahn, S. Boyault, J. E. Brock, A. Broeks, A. Butler, C. Desmedt, L. Dirix, S. Dronov, A. Fatima, J. A. Foekens, M. Gerstung, G. K. J. Hooijer, S. J. Jang, D. R. Jones, H.-Y. Kim, T. A. King, S. Krishnamurthy, H. J. Lee, J.-Y. Lee, Y. Li, S. McLaren, A. Menzies, V. Mustonen, S. O'Meara, I. Pauporté, X. Pivot, C. A. Purdie, K. Raine, K. Ramakrishnan, F. G. Rodríguez-González, G. Romieu, A. M. Sieuwerts, P. T. Simpson, R. Shepherd, L. Stebbings, O. A. Stefansson, J. Teague, S. Tommasi, I. Treilleux, G. G. Van den Eynden, P. Vermeulen, A. Vincent-Salomon, L. Yates, C. Caldas, L. van't Veer, A. Tutt, S. Knappskog, B. K. T. Tan, J. Jonkers, Å. Borg, N. T. Ueno, C. Sotiriou, A. Viari, P. A. Futreal, P. J. Campbell, P. N. Span, S. Van Laere, S. R. Lakhani, J. E. Eyfjord, A. M. Thompson, E. Birney, H. G. Stunnenberg, M. J. van de Vijver, J. W. M. Martens, A.-L. Børresen-Dale, A. L. Richardson, G. Kong, G. Thomas, M. R. Stratton, Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*. **534**, 47–54 (2016).
32. S. Paik, S. Shak, G. Tang, C. Kim, J. Baker, M. Cronin, F. L. Baehner, M. G. Walker, D. Watson, T. Park, W. Hiller, E. R. Fisher, D. L. Wickerham, J. Bryant, N. Wolmark, A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **351**, 2817–2826 (2004).
33. L. J. Solin, R. Gray, F. L. Baehner, S. M. Butler, L. L. Hughes, C. Yoshizawa, D. B. Cherbavaz, S. Shak, D. L. Page, G. W. Sledge Jr, N. E. Davidson, J. N. Ingle, E. A. Perez, W. C. Wood, J. A. Sparano, S. Badve, A multigene expression assay to predict local recurrence risk for ductal carcinoma in situ of the breast. *J. Natl. Cancer Inst.* **105**, 701–710 (2013).
34. N. Poulakaki, G.-M. Makris, A.-M. Papanota, F. Marineli, A. Marinelis, M.-J. Battista, D. Boehm, A. Psyrri, T. N. Sergentanis, Ki-67 Expression as a Factor Predicting Recurrence of Ductal Carcinoma In Situ of the Breast: A Systematic Review and Meta-Analysis. *Clin. Breast Cancer*. **18**, 157–167.e6 (2018).
35. E. de Azambuja, F. Cardoso, G. de Castro Jr, M. Colozza, M. S. Mano, V. Durbecq, C. Sotiriou, D. Larsimont, M. J. Piccart-Gebhart, M. Paesmans, Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. *Br. J. Cancer*. **96**, 1504–1513 (2007).
36. L.-N. Yu, Z. Liu, Y. Tian, P.-P. Zhao, X. Hua, FAP-a and GOLPH3 Are Hallmarks of DCIS Progression to Invasive Breast Cancer. *Front. Oncol.* **9**, 1424 (2019).
37. R. Kalluri, The biology and function of fibroblasts in cancer. *Nat. Rev. Cancer*. **16**, 582–598 (2016).
38. A. Costa, Y. Kieffer, A. Scholer-Dahirel, F. Pelon, B. Bourachot, M. Cardon, P. Sirven, I. Magagna, L. Fuhrmann, C. Bernard, C. Bonneau, M. Kondratova, I. Kuperstein, A. Zinovyev, A.-M. Givel, M.-C. Parrini, V. Soumelis, A. Vincent-



- Salomon, F. Mehta-Grigoriou, Fibroblast Heterogeneity and Immunosuppressive Environment in Human Breast Cancer. *Cancer Cell*. **33**, 463–479.e10 (2018).
39. A. C. Nelson, H. L. Machado, K. L. Schwertfeger, Breaking through to the Other Side: Microenvironment Contributions to DCIS Initiation and Progression. *J. Mammary Gland Biol. Neoplasia*. **23**, 207–221 (2018).
40. K. L. Schwertfeger, W. Xian, A. M. Kaplan, S. H. Burnett, D. A. Cohen, J. M. Rosen, A Critical Role for the Inflammatory Response in a Mouse Model of Preneoplastic Progression. *Cancer Research*. **66** (2006), pp. 5676–5685.
41. A. C. Vargas, A. E. McCart Reed, N. Waddell, A. Lane, L. E. Reid, C. E. Smart, S. Cocciardi, L. da Silva, S. Song, G. Chenevix-Trench, P. T. Simpson, S. R. Lakhani, Gene expression profiling of tumour epithelial and stromal compartments during breast cancer progression. *Breast Cancer Res. Treat.* **135**, 153–165 (2012).
42. C. F. Cowell, B. Weigelt, R. A. Sakr, C. K. Y. Ng, J. Hicks, T. A. King, J. S. Reis-Filho, Progression from ductal carcinoma in situ to invasive breast cancer: Revisited. *Mol. Oncol.* **7**, 859–869 (2013).
43. D. Juric, P. Castel, M. Griffith, O. L. Griffith, H. H. Won, H. Ellis, S. H. Ebbesen, B. J. Ainscough, A. Ramu, G. Iyer, R. H. Shah, T. Huynh, M. Mino-Kenudson, D. Sgroi, S. Isakoff, A. Thabet, L. Elamine, D. B. Solit, S. W. Lowe, C. Quadt, M. Peters, A. Derti, R. Schegel, A. Huang, E. R. Mardis, M. F. Berger, J. Baselga, M. Scaltriti, Convergent loss of PTEN leads to clinical resistance to a PI(3)K $\alpha$  inhibitor. *Nature*. **518**, 240–244 (2015).
44. M. Kim, K. Jang, P. Miller, M. Picon-Ruiz, T. M. Yeasky, D. El-Ashry, J. M. Slingerland, VEGFA links self-renewal and metastasis by inducing Sox2 to repress miR-452, driving Slug. *Oncogene*. **36**, 5199–5211 (2017).
45. H. Tong, J.-Q. Ke, F.-Z. Jiang, X.-J. Wang, F.-Y. Wang, Y.-R. Li, W. Lu, X.-P. Wan, Tumor-associated macrophage-derived CXCL8 could induce ER $\alpha$  suppression via HOXB13 in endometrial cancer. *Cancer Lett.* **376**, 127–136 (2016).
46. S. A. Hussain, R. Ganesan, G. Reynolds, L. Gross, A. Stevens, J. Pastorek, P. G. Murray, B. Perunovic, M. S. Anwar, L. Billingham, N. D. James, D. Spooner, C. J. Poole, D. W. Rea, D. H. Palmer, Hypoxia-regulated carbonic anhydrase IX expression is associated with poor survival in patients with invasive breast cancer. *Br. J. Cancer*. **96**, 104–109 (2007).
47. D. Jones, E. R. Pereira, T. P. Padera, Growth and Immune Evasion of Lymph Node Metastasis. *Front. Oncol.* **8**, 36 (2018).
48. M. Greaves, C. C. Maley, Clonal evolution in cancer. *Nature*. **481**, 306–313 (2012).
49. S. Vickovic, G. Eraslan, J. Klughammer, L. Stenbeck, F. Salmen, T. Aijo, R. Bonneau, L. Bergenstraahle, J. Gould, M. Ronaghi, J. Frisen, J. Lundeberg, A.

- Regev, P. L. Staahl, High-density spatial transcriptomics arrays for in situ tissue profiling. *bioRxiv* (2019), p. 563338.
50. T. Krzywkowski, M. Kühnemund, M. Nilsson, Chimeric padlock and iLock probes for increased efficiency of targeted RNA detection. *RNA*. **25**, 82–89 (2019).
  51. D. Gyllborg, C. M. Langseth, X. Qian, E. Choi, S. M. Salas, M. M. Hilscher, E. S. Lein, M. Nilsson, Hybridization-based in situ sequencing (HybISS) for spatially resolved transcriptomics in human and mouse brain tissue. *Nucleic Acids Res.* **48**, e112 (2020).
  52. N. Bolli, H. Avet-Loiseau, D. C. Wedge, P. Van Loo, L. B. Alexandrov, I. Martincorena, K. J. Dawson, F. Iorio, S. Nik-Zainal, G. R. Bignell, J. W. Hinton, Y. Li, J. M. C. Tubio, S. McLaren, S. O’Meara, A. P. Butler, J. W. Teague, L. Mudie, E. Anderson, N. Rashid, Y.-T. Tai, M. A. Shammash, A. S. Sperling, M. Fulciniti, P. G. Richardson, G. Parmigiani, F. Magrangeas, S. Minvielle, P. Moreau, M. Attal, T. Facon, P. A. Futreal, K. C. Anderson, P. J. Campbell, N. C. Munshi, Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat. Commun.* **5**, 2997 (2014).
  53. I. Kozarewa, Z. Ning, M. A. Quail, M. J. Sanders, Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nature* (2009) (available at <https://www.nature.com/nmeth/journal/v6/n4/abs/nmeth.1311.html>).
  54. J. Salvatier, T. V. Wiecki, C. Fonnesbeck, Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.* **2**, e55 (2016).
  55. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* **25**, 1754–1760 (2009).
  56. S. C. Dentro, D. C. Wedge, P. Van Loo, Principles of Reconstructing the Subclonal Architecture of Cancers. *Cold Spring Harb. Perspect. Med.* **7** (2017), doi:10.1101/cshperspect.a026625.
  57. J. C. Caicedo, A. Goodman, K. W. Karhohs, B. A. Cimini, J. Ackerman, M. Haghghi, C. Heng, T. Becker, M. Doan, C. McQuin, M. Rohban, S. Singh, A. E. Carpenter, Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. *Nat. Methods.* **16**, 1247–1253 (2019).
  58. S. Nik-Zainal, L. B. Alexandrov, D. C. Wedge, P. Van Loo, C. D. Greenman, K. Raine, D. Jones, J. Hinton, J. Marshall, L. A. Stebbings, A. Menzies, S. Martin, K. Leung, L. Chen, C. Leroy, M. Ramakrishna, R. Rance, K. W. Lau, L. J. Mudie, I. Varela, D. J. McBride, G. R. Bignell, S. L. Cooke, A. Shlien, J. Gamble, I. Whitmore, M. Maddison, P. S. Tarpey, H. R. Davies, E. Papaemmanuil, P. J. Stephens, S. McLaren, A. P. Butler, J. W. Teague, G. Jönsson, J. E. Garber, D. Silver, P. Miron, A. Fatima, S. Boyault, A. Langerød, A. Tutt, J. W. M. Martens, Samuel A J, Å. Borg, A. V. Salomon, G. Thomas, A.-L. Børresen-Dale, A. L. Richardson, M. S.

- Neuberger, P. Andrew Futreal, P. J. Campbell, M. R. Stratton, Mutational Processes Molding the Genomes of 21 Breast Cancers. *Cell*. **149**, 979–993 (2012).
59. A. C. Ruifrok, D. A. Johnston, Quantification of histochemical staining by color deconvolution. *Anal. Quant. Cytol. Histol.* **23**, 291–299 (2001).
60. BRADSKI, G, The OpenCV library. *Dr Dobb's J. Software Tools*. **25**, 120–125 (2000).
61. S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, scikit-image contributors, scikit-image: image processing in Python. *PeerJ*. **2**, e453 (2014).
62. D. Hörl, F. Rojas Rusak, F. Preusser, P. Tillberg, N. Randel, R. K. Chhetri, A. Cardona, P. J. Keller, H. Harz, H. Leonhardt, M. Treier, S. Preibisch, BigStitcher: reconstructing high-resolution image datasets of cleared and expanded samples. *Nat. Methods*. **16**, 870–874 (2019).
63. J. H. Taube, J. I. Herschkowitz, K. Komurov, A. Y. Zhou, S. Gupta, J. Yang, K. Hartwell, T. T. Onder, P. B. Gupta, K. W. Evans, B. G. Hollier, P. T. Ram, E. S. Lander, J. M. Rosen, R. A. Weinberg, S. A. Mani, Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 15449–15454 (2010).
64. P. Savas, R. Salgado, C. Denkert, C. Sotiriou, P. K. Darcy, M. J. Smyth, S. Loi, Clinical relevance of host immunity in breast cancer: from TILs to the clinic. *Nat. Rev. Clin. Oncol.* **13**, 228–241 (2016).
65. I. Arganda-Carreras, C. O. S. Sorzano, J. Kybic, C. Ortiz-de Solorzano, in *Second ImageJ User & Developer Conference* (2008; [https://www.researchgate.net/profile/Carlos\\_Sorzano/publication/267257572\\_bUnwarpJ\\_Consistent\\_and\\_Elastic\\_Registration\\_in\\_ImageJ\\_Methods\\_and\\_Applications/inks/548a8d730cf214269f1ac7ee.pdf](https://www.researchgate.net/profile/Carlos_Sorzano/publication/267257572_bUnwarpJ_Consistent_and_Elastic_Registration_in_ImageJ_Methods_and_Applications/inks/548a8d730cf214269f1ac7ee.pdf)), vol. 12.