Cold Spring Harbor Perspectives in Medicine
www.perspectivesinmedicine.org

Gene-Expression Profiling to Decipher Breast Cancer Inter- and Intratumor Heterogeneity

Alexander Swarbrick,[1,2] Aranzazu Fernandez-Martinez,[3,4] and Charles M. Perou[3,4]

[1]Cancer Ecosystems Program, Garvan Institute of Medical Research, Darlinghurst, New South Wales 2010, Australia

[2]St Vincent's Clinical School, Faculty of Medicine, University of New South Wales, Sydney, New South Wales 2052, Australia

[3]Lineberger Comprehensive Center, University of North Carolina, Chapel Hill, North Carolina 27599, USA

[4]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27514, USA

Correspondence: a.swarbrick@garvan.org.au

Breast cancer is heterogeneous and differs substantially across different tumors (intertumor heterogeneity) and even within an individual tumor (intratumor heterogeneity). Gene-expression profiling has considerably impacted our understanding of breast cancer biology. Four main "intrinsic subtypes" of breast cancer (i.e., luminal A, luminal B, HER2-enriched, and basal-like) have been consistently identified by gene expression, showing significant prognostic and predictive value in multiple clinical scenarios. Thanks to the molecular profiling of breast tumors, breast cancer is a paradigm of treatment personalization. Several standardized prognostic gene-expression assays are presently being used in the clinic to guide treatment decisions. Moreover, the development of single-cell-level resolution molecular profiling has allowed us to appreciate that breast cancer is also heterogeneous within a single tumor. There is an evident functional heterogeneity within the neoplastic and tumor microenvironment cells. Finally, emerging insights from these studies suggest a substantial cellular organization of neoplastic and tumor microenvironment cells, thus defining breast cancer ecosystems and highlighting the importance of spatial localizations.

Breast cancer has long been recognized as a heterogeneous disease. For many decades and to this day, breast cancers have been pathologically diagnosed as estrogen receptor (ER)-positive/-negative, human epidermal receptor 2 (HER2)-positive/-negative, and progesterone receptor (PR)-positive/-negative. These three therapeutic biomarkers then, in essence, define three clinical groups that are ER/PR$^+$/HER2$^-$, HER2$^+$, and triple-negative breast cancer (TNBC) (i.e., negative for ER, PR, and HER2). This classification provides not only prognostic information but also treatment directions, which is the main driving reason for ER, PR, and HER2 testing. With the development of DNA microarray technology for profiling transcriptomes, breast cancers have been additionally classified into four main genomic "intrinsic" subtypes: luminal A, luminal B,

Editors: Jane E. Visvader, Jeffrey M. Rosen, and Samuel Aparicio
Additional Perspectives on Breast Cancer: From Fundamental Biology to Therapeutic Strategies available at www.perspectivesinmedicine.org

Advanced Online Article. Cite this article as *Cold Spring Harb Perspect Med* doi: 10.1101/cshperspect.a041320

Cold Spring Harbor Perspectives in Medicine
www.perspectivesinmedicine.org

HER2-enriched, and basal-like (Perou et al. 2000; Sørlie et al. 2001). Molecular profiling of hormone receptor (HR)-positive (i.e., ER- and/or PR-positive), HER2-negative breast cancers have provided several standardized prognostic gene-expression signatures that are being used in clinical practice, all of which provide additional information for prognostication and prediction of treatment benefit. With the development of single-cell-level resolution molecular profiling approaches, it is now recognized that breast cancers are heterogeneous not only among different tumors but also within a single tumor. There is an evident functional heterogeneity within the neoplastic cell population in breast cancers. In addition, there is growing awareness of the importance of the tumor microenvironment (TME), which includes host stromal cells and immune cells, beyond the neoplastic cell-intrinsic heterogeneity. In this review, we will discuss breast cancer heterogeneity, both intertumoral and intratumoral, based on gene-expression and related "omics" approaches, focusing on insights for human disease. We will first discuss studies analyzing tissue bulk-based approaches and then those using single-cell and spatially resolved techniques. The emerging insights from these studies suggest a substantial cellular organization of neoplastic and TME cells, thus defining breast cancer ecosystems.

DNA MICROARRAYS TO NEXT-GENERATION SEQUENCING FOR GENE EXPRESSION

Gene-expression profiling with DNA microarrays and high-throughput sequencing technologies like RNA sequencing (RNA-seq) allows for the identification of tens of thousands of transcripts in a single experiment. Thus, we can now identify gene-expression signatures, defined as groups of genes with a unique and characteristic expression pattern, representing specific cell types and biological pathways (Massagué 2007; Sotiriou and Pusztai 2009; Fan et al. 2011) instead of focusing on single gene candidates. These signatures usually have a biologic, prognostic, or predictive value, and some of them are currently commercially available and commonly used for breast cancer management.

The first signature identification approaches in breast cancer were based on clustering methods. These experiments allowed for the classification of breast tumors into luminal A, luminal B, HER2-enriched, and basal-like as four main molecular entities, or intrinsic subtypes, with different tumor biology and prognostic impact (Perou et al. 2000; Sørlie et al. 2001). The first generation of clinically available gene-expression prognostic signatures (i.e., MammaPrint and Oncotype DX) was created by genomically comparing low- versus high-risk breast tumors. Over time, multiple gene-expression signatures have been developed and validated to track different molecular pathways and specific biological processes (i.e., proliferation, hypoxia, cell differentiation, immune-cell states), and some of these signatures also predict sensitivity to treatments (Gatza et al. 2010; Fan et al. 2011; Iglesia et al. 2014; Newman et al. 2015; Hollern et al. 2019; Garcia-Recio et al. 2020). Moreover, the integration of gene-expression signatures and clinicopathologic factors has proven to significantly increase the prognostic ability compared to either one alone (Fan et al. 2011); and the integration of gene signatures, clinicopathologic features, DNA copy number alterations (CNAs), and DNA somatic mutations can predict response to treatment with improved accuracy over clinical factors alone (Tanioka et al. 2018).

For the first 15 yr of gene-expression profiling, hybridization-based platforms like DNA microarrays were the gold standard technology for gene-expression analysis. However, with the advent of RNA-seq based on next-generation sequencing technologies, a new gold standard has been set that has clear advantages when compared to other gene-expression profiling platforms (Table 1). In recent years, RNA-seq has largely replaced microarrays for the discovery of breast cancer prognostic and predictive biomarkers.

Hybridization platforms are high throughput and relatively inexpensive. However, they have several limitations (Wang et al. 2009):

- They can analyze only pre-defined sequences,
- Their dynamic range is limited,

Table 1. Main characters of DNA microarrays and RNA sequencing (RNA-seq)

	Microarrays	RNA-seq
Number of genes	High throughput >20,000	High throughput Whole transcriptome
Sample preparation	RNA extraction Reverse transcribe sample	RNA extraction Reverse transcribe sample Fragmentation Library preparation
Processing steps	Label cDNAs Hybridization to array	Sequencing
Instrument	Microarray scanner	Sequencer
Time on sample processing	17–18 h (overnight)	24 h
Time on data processing	Short	Long
Cost	Moderate	Expensive

- The hybridization can be nonspecific (i.e., cross-hybridization) (Okoniewski and Miller 2006), and
- They usually show highly variable results for lowly expressed genes.

RNA-seq is technically a significant improvement beyond hybridization-based approaches in many ways. Compared to DNA microarrays, RNA-seq has several advantages (Wang et al. 2009):

- No prior sequence information is required,
- Allows for a precise and more quantitative measure of the gene expression,
- High sensitivity for the detection of low-abundance transcripts,
- Wider dynamic range,
- Can reveal sequence variations (i.e., RNA-seq), thus facilitating the discovery of alternative splicing variants, structural variants, and single-nucleotide variations (SNVs), and
- Greater reproducibility.

Due to all these advantages and the fact that RNA-seq provides the entire transcriptome to be surveyed in a high-throughput and quantitative manner, it has become the current gold standard laboratory tool for gene-expression profiling. However, this technology also comes with several challenges, including:

- RNA-seq is computationally challenging and requires the constant development of new methods to store, retrieve, process, and analyze a large amount of data,
- The lack of standardization among sequencing laboratories and read depths can compromise reproducibility, and
- Although the cost of RNA-seq has decreased in recent years, it is relatively high.

PROGNOSTIC SIGNATURES IN BREAST CANCER

With these robust technologies in hand, be they DNA microarrays or RNA-seq, breast cancer biologists used these tools to identify gene-expression patterns that might be of clinical usefulness. During the last two decades, gene-expression platforms have significantly increased our ability to estimate distant recurrence of early-stage HR-positive/HER2-negative breast cancers. Comparing gene-expression profiles of high-risk and low-risk early-stage HR-positive/HER2-negative breast tumors yielded different prognostic gene-expression signatures. Some of these are currently commercially available, endorsed by clinical guidelines, and routinely used in clinical practice. These signatures include Oncotype DX Recurrence Score (RS), MammaPrint 70-gene signature, Prosigna risk of recurrence (ROR) score, EndoPredict, and Breast Cancer Index (Table 2; Cardoso et al. 2019; Andre et al. 2022).

Oncotype DX (Genomic Health, Redwood, CA, USA) assesses the expression of 21 genes (16 breast cancer–related genes and five reference genes) using reverse transcription polymerase chain reaction (RT-PCR), providing a risk score (RS) that classifies the patient in low-, intermediate-, and high-risk prognostic groups (Paik et al. 2004). The RS prognostic value was first retrospectively evaluated in two clinical trials, NSABP (Paik et al. 2004) and TransATAC (Dowsett et al. 2010). The RS predictive value for chemotherapy benefit was also studied in two retrospective studies: NSABP B-20 (Paik et al. 2006) and SWOG-8814 (Albain et al. 2010). According to the ASCO guidelines, NSABP B-20 data were confounded by the data set initially used to generate the assay, and the SWOG-8814 data was hypothesis-generating due to the small sample set and no additional prediction beyond 5 years. The prognostic ability of Oncotype DX RS has been analyzed in the prospective clinical trial TAILORx (Sparano et al. 2018), where the RS cutoffs between different prognostic groups were changed for this purpose. Patients with RS $\leq$ 10 were considered low risk and were assigned not to receive chemotherapy, patients with RS $\geq$ 26 were assigned to receive chemotherapy, and patients with RS between 11 and 25 were randomly assigned to not receive or receive adjuvant chemotherapy. This clinical trial demonstrated the noninferiority of endocrine therapy alone compared to a chemotherapy and endocrine therapy combination in patients with an RS score between 16 and 25, with a small benefit from chemotherapy in patients 50 years old and younger. Data from the RxPONDER trial suggests that the benefit of adjuvant chemotherapy is small in patients with RS < 25 and 1–3 positive nodes (Kalinsky et al. 2021). However, RS failed to predict chemotherapy benefit (i.e., statistically nonsignificant interaction test). In addition, premenopausal patients younger than 50 again showed a large absolute benefit (i.e., an absolute difference of 4.9 percentage points) from adjuvant chemotherapy. Whether the observed chemotherapy benefits in these studies is directly or indirectly from chemotherapy-induced ovarian suppression remains unknown (Sparano et al. 2018; Kalinsky et al. 2021) and thus remains a pressing question to be addressed. Oncotype DX RS, therefore, provides relevant prognostic information and is currently recommended in clinical guidelines with the strongest level of evidence (Cardoso et al. 2019; Andre et al. 2022).

MammaPrint (Agendia, Amsterdam, The Netherlands) is a 70-gene signature that uses the microarray technology that required initially sending fresh frozen breast tumor samples to a central laboratory, and then switched to formalin-fixed paraffin-embedded (FFPE). The prognostic value of this assay has been validated in multiple retrospective (Van't Veer et al. 2002; Buyse et al. 2006) and prospective (Drukker et al. 2013; Cardoso et al. 2016) studies. Similar to Oncotype DX, MammaPrint has not proven to have predictive value for chemotherapy benefit. MammaPrint is also included in most breast cancer guidelines with a high level of evidence (Cardoso et al. 2019; Andre et al. 2022).

Prosigna (Veracyte, San Francisco, CA, USA) is a prognostic gene-expression assay that uses 58 genes (PAM50 signature) along with clinicopathological features to calculate the ROR score, a prognostic index that predicts the probability of cancer recurrence. Prosigna is a decentralized test that can be run using FFPE breast tumor samples. Apart from the ROR score (and outside the United States), Prosigna also provides the breast cancer–intrinsic subtype information. Retrospectively validated in two different clinical trials (Dowsett et al. 2010; Gnant et al. 2014), Prosigna can be used as a prognostic tool to guide the use of adjuvant therapy (Cardoso et al. 2019; Andre et al. 2022) and, unlike Oncotype DX and MammaPrint, the ROR score also has a prognostic value for late recurrences that might occur between 5 and 10 years postdiagnosis.

The Breast Cancer Index (Biotheranostics, San Diego, CA, USA) is a gene-expression signature that includes the *HOXB13/IL17BR* ratio and the molecular grade index. The prognostic score combines both pieces of information and provides an individual risk assessment for overall and late distant recurrence (Sgroi et al. 2013a, b). The predictive component is based on the *HOXB13/IL17BR* ratio and provides a classification (i.e., high vs. low) for the benefit of extended

Table 2. Characteristics of the prognostic gene-expression assays in early-stage HR-positive, HER2-negative breast cancer

Signature	Oncotype DX	MammaPrint	Prosigna	Breast Cancer Index	EndoPredict
Provider	Genomic Health	Agendia	Veracyte	Biotheranostics	Myriad Genetics
Tissue for analysis[a]	FFPE	FFPE	FFPE	FFPE	FFPE
Platform[b]	RT-PCR	Microarray	nCounter	RT-PCR	RT-PCR
Number of genes	21 genes	70 genes	58 genes	11 genes	12 genes
Central testing	Yes	Yes	No	Yes	No
Risk stratification	Three groups	Two groups	Three groups	Two groups	Two groups
Indications	Pre- and postmenopausal N0 Postmenopausal 1–3 positive nodes	Pre- and postmenopausal N0 Postmenopausal 1–3 positive nodes	Postmenopausal N0	Postmenopausal N0 Postmenopausal 1–3 positive nodes Prediction of benefit from extended (>5 yr) endocrine therapy	Postmenopausal N0 Postmenopausal 1–3 positive nodes

[a](FFPE) Formalin-fixed paraffin-embedded.
[b](RT-PCR) Reverse transcription polymerase chain reaction.

endocrine therapy (Bartlett et al. 2016). Unlike other assays, the Breast Cancer Index can be used to guide an extension of the endocrine treatment in patients with node-negative or node-positive (one to three nodes only) disease beyond 5 years of treatment (Andre et al. 2022).

EndoPredict (Myriad Genetics, Salt Lake City, UT, USA) is a prognostic gene-expression signature of 12 genes (eight cancer-related genes, three reference genes, and one DNA control gene). The commercialized test (EPclin) also considers clinical factors (tumor size and node status) to inform about the risk of recurrence and indication of adjuvant chemotherapy. This assay has been retrospectively tested in the ABCSG6, the ABCSG8, and the TransATAC clinical trials (Dubsky et al. 2013; Buus et al. 2016), and the European and American guidelines recommend it as a prognostic tool for treatment decisions (Cardoso et al. 2019; Andre et al. 2022).

These genomic assays help clinicians and patients to decide the need for adjuvant chemotherapy and the extension of hormone therapy to treat early-stage HR-positive/HER2-negative breast cancer. Despite multiple efforts in the molecular characterization of HER2-positive breast cancer and TNBC, gene-expression assays are not yet recommended by clinical guidelines to guide treatment decisions within these two clinical subtypes of breast cancer.

MOLECULAR CHARACTERIZATION OF BREAST CANCER BY TCGA

The Cancer Genome Atlas (TCGA) revolutionized the study of human cancers in two ways: first, it applied a set of at least six different "omic" technologies onto a common set of tumors (i.e., whole-exome sequencing, DNA–single-nucleotide polymorphism (SNP) microarrays, DNA methylation microarrays, gene expression by DNA microarrays and/or RNA-seq, miRNA sequencing, and reverse-phase protein arrays [RPPAs]), thus providing an unprecedented "multiplatform" data set. Second, all these data were made immediately public, and have literally been used in thousands of publications across tens of countries. Breast cancers were highly represented among the TCGA tumor types, which in the end represented 1100/10,000 cancers profiled (Hoadley et al. 2018). Multiple TCGA Breast Cancer Studies were published, including the most highly cited TCGA-sponsored paper from 2012, one of the first multiplatform analyses of a human tumor type (Cancer Genome Atlas Network 2012). A highly novel computational approach was employed in this paper where a single classification of breast cancers was made via the integrated analysis of all six data types at once, as opposed to most previous methods like those described above, where classifications are made based upon a single technology. This method, called "clustering of cluster assignments" (COCA), identified four main groups of breast tumors, which were highly similar to the four major breast cancer intrinsic subtypes based upon gene-expression patterns alone (Fig. 1); note that in the COCA method, each of the six omics data types was equally weighted, thus demonstrating that the gene-expression data well represents all these data and that these gene-expression-defined intrinsic subtypes are also highly correlated with DNA and protein features, which also define similar subtypes.

Luminal A is the most common intrinsic subtype among all breast cancers and within HR-positive/HER2-negative breast cancer. These tumors have heterogeneous CNAs and somatic mutations (Ciriello et al. 2013). Luminal A tumors have the highest number of recurrently mutated genes but the smallest number of mutations per tumor. Among the most frequently mutated genes are *PIK3CA* (45%), followed by *MAP3K1*, *GATA3*, *TP53*, *CDH1*, and *MAP2K4*. This subtype is characterized by typically high expression of ER and PR (both mRNA and protein levels) and an increased expression of the luminal signature, which contains genes that are also highly expressed in the mammary ducts' luminal epithelium (*ESR1*, *GATA3*, *FOXA1*, *XBP1*, and *MYB*). Compared to luminal B, luminal A tumors usually present lower proliferation rates (Cancer Genome Atlas Network 2012).

Luminal B breast cancer is similar and related to luminal A but is also unique in the profile of DNA CNAs and somatic mutations, and often shows a DNA hypermethylated phenotype.

Cold Spring Harbor Perspectives in Medicine
www.perspectivesinmedicine.org

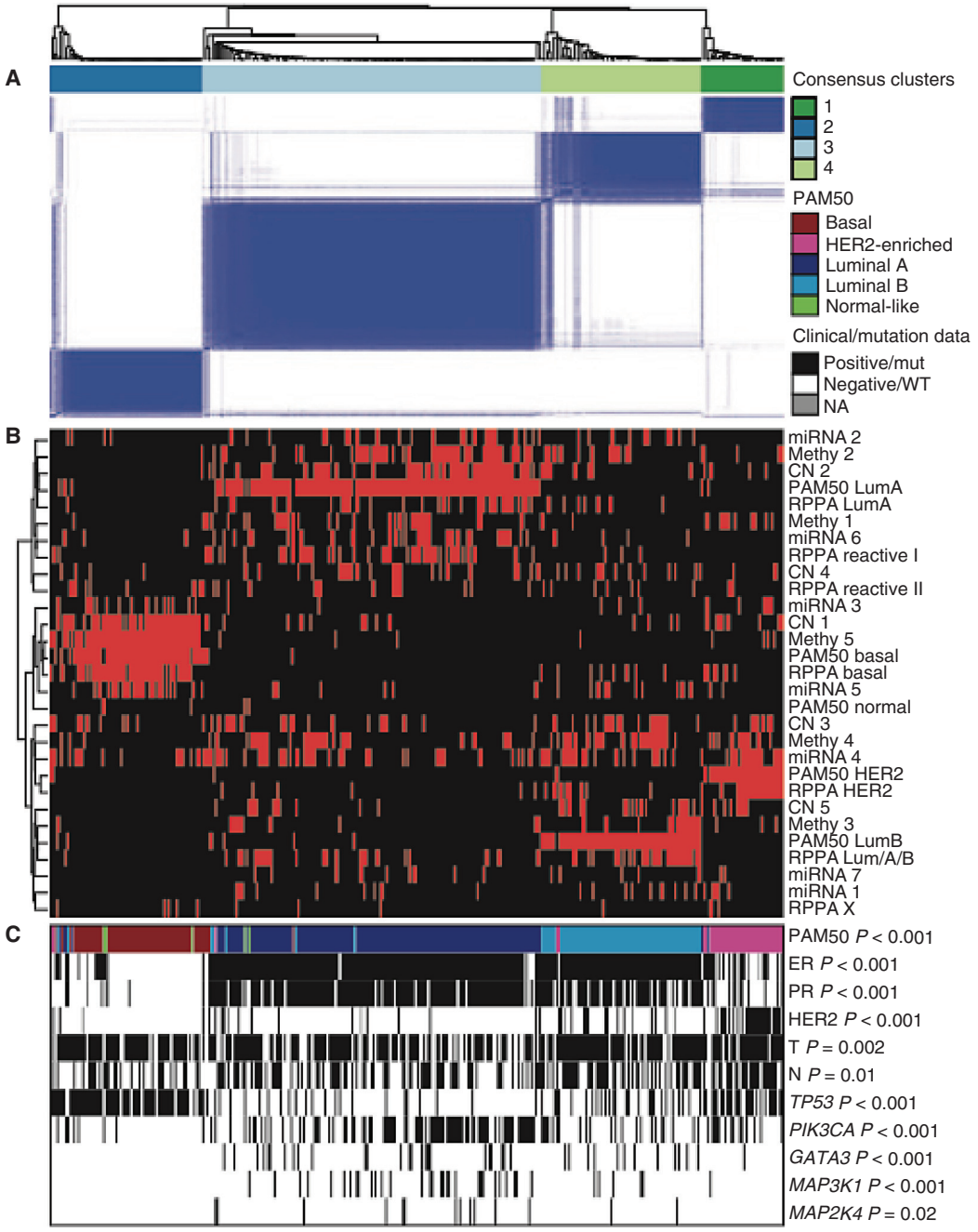

Figure 1. Coordinated analysis of breast cancer subtypes defined from five different genomic/proteomic platforms. (*A*) Consensus clustering analysis of the subtypes identifies four major groups (samples, $n = 348$). The blue and white heat map displays sample consensus. (*B*) Heat-map display of the subtypes defined independently by miRNAs, DNA methylation, copy number (CN), PAM50 mRNA expression, and reverse-phase protein array (RPPA) expression. The red bar indicates membership of a cluster type. (*C*) Associations with molecular and clinical features. *P* values were calculated using a chi-squared test. (Figure is reprinted from the Cancer Genome Atlas Network (2012) with permission from the author, C.M. Perou.)

Compared to luminal A tumors, luminal B tumors have a lower frequency of *PIK3CA* mutations and a higher frequency of *TP53* mutations. Luminal B tumors tend to have a lower expression of luminal signature, lower expression of PR (protein and mRNA), and a higher proliferation rate than luminal A tumors (Cancer Genome Atlas Network 2012).

HER2-Enriched is the predominant genomic subtype within pathological HER2-positive tumors, but not all HER2-positive tumors are HER2-enriched, and not all clinically defined $HER2^+$ tumors are of the HER2-enriched gene-expression phenotype. Clinically defined HER2-positive tumors are probably the most heterogeneous breast cancer clinical type among those analyzed by TCGA, where ~50% of HER2-positive tumors are HER2-enriched. An HER2-enriched subtype is mainly characterized by DNA amplification of the HER2 gene (*ERBB2*) and a high frequency of *TP53* mutation (~75%). Tumors that are both HER2-enriched and HER2-positive show a significantly higher expression of several tyrosine kinase receptors, including FGFR4, EGFR, HER2 itself, and genes within the *ERBB2* amplicon, including GRB7 and sometimes TOP2A. HER2-enriched tumors usually have a high frequency of *PIK3CA* mutations (39%), a lower frequency of *PTEN* mutations, high aneuploidy, and the highest somatic mutation rate (Cancer Genome Atlas Network 2012).

Basal-like is the predominant genomic subtype within TNBC, with ~75% of all TNBCs being basal-like. Basal-like breast tumors harbor a high frequency of *TP53* mutations (80%), and most of these tumors have lost the TP53 function based on the combination of the *TP53* mutation frequency and the inferred TP53 pathway activity. Although *PIK3CA* is the second-most-mutated gene in basal-like breast tumors, its frequency is relatively low (~9%) compared to *TP53* and much lower than that in luminal tumors (~40%); however, despite this low *PIK3CA* mutation rate, the inferred PI3K pathway activity based on RPPA proteomics is the highest in this subtype (Cancer Genome Atlas Network 2012). At the gene-expression level, basal-like tumors are characterized by high expressions of keratins 5, 6, and 17, low expression of *RB1* (which also shows a 5% mutation rate), and increased expression of proliferation genes. Hyperactivation of MYC and H1F1 pathways are also frequent. Basal-like tumors usually have the lowest level of global DNA methylation and are confirmed to have a strong association with germline *BRCA1* mutations (Cancer Genome Atlas Network 2012).

Last, several other TCGA Breast Cancer publications should be highlighted, including a multiplatform molecular description of lobular cancers (Ciriello et al. 2015), a comparison of omics features between breast cancer patients of African versus European ancestry (Huo et al. 2017), and a thorough comparison of multiple pathology features including grade versus all the genomic data (Heng et al. 2017). Very recently, a complete analysis of all 1100 TCGA breast cancers and all histologies, including a focus on six rare subtypes (cribriform, micropapillary, mucinous, papillary, metaplastic, and invasive ductal carcinoma [IDC] with medullary pattern), was completed, and, importantly, where histology and genomics were linked together to arrive at 12 consensus groups (Thennavan et al. 2021); these two consensus groups, highlighted in Figure 2, link together histology and genomics into a single framework that should improve our ability to understand the complex biology of each breast cancer clinical study.

In addition to TCGA, multiple international efforts have helped decipher breast cancer biology using a multi-omics approach. Thus, Metabric (Molecular Taxonomy of Breast Cancer International Consortium) used an integrated clustering approach of genome and transcriptome data from 2000 breast tumors describing 10 integrative clusters with distinct clinical outcomes (Curtis et al. 2012). Other international efforts, like the International Cancer Genome Consortium (ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium 2020), the Broad Institute (Banerji et al. 2012), and the National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (Krug et al. 2020), have also contributed to a better understanding of the molecular mechanisms of breast tumors. More recently, the Sweden Cancerome Analysis Network—Breast (SCAN-B) consortium was created as a multicen-

Cold Spring Harbor Perspectives in Medicine
www.perspectivesinmedicine.org

Figure 2. The Cancer Genome Atlas (TCGA) breast cancer classification based upon molecular and histologic features combined. Schematic representation of 12 consensus groups defined by histology and gene-expression analyses of the TCGA-BRCA data set and organized by differentiation score. These groups are connected by an outer ring based on differentiation score (D score: lowest differentiation to highest differentiation arranged in an anticlockwise direction). From the D score ring inward: the second ring exhibits PAM50 subtype association with D score (red: basal, pink: HER2E, blue: LumA and LumB, yellow: claudin low). The next ring highlights the proliferation gene signature, which is high in all biological groups with a low D score but also in one group with a high D score, namely, IDC-LumB. The next two rings represent the descending abundance of copy number alteration (CNA) events (4p loss, 2p gain) and mutation events (Tp53 mutation) associated with ascending D score. The final ring exhibits decreasing immunological gene signatures in relation to ascending D score. The innermost pie chart exhibits the clinical immunohistochemistry (IHC) status found in these 12 breast cancer consensus groups.

ter prospective study to analyze breast tumors with next-generation technologies to develop and validate new biomarker assays to better personalize cancer treatment (Saal et al. 2015). Thanks to this initiative, gene expression and clinical data from a cohort of 7743 patients with breast cancer have been analyzed (Staaf et al. 2022). Genomic and clinical data from these international consortiums is usually shared with the scientific community, providing the opportunity to accelerate breast cancer research.

SINGLE-CELL AND SPATIALLY RESOLVED GENE-EXPRESSION ANALYSIS

Breast cancers typically exhibit extensive immune cells and stromal cells infiltration, along with functional heterogeneity among neoplastic cells. There is a growing awareness of the importance of the TME, beyond neoplastic cell-intrinsic heterogeneity, to the role of the TME in breast cancer etiology, response to therapy, and patient survival outcomes. Gene-expression analyses of bulk tissues have led to paradigm shifts in our understanding of breast cancer biology. However, a limitation of such approaches is that they only measure tissue-average gene expression, and thus it is often challenging to know what gene(s) are being expressed in what cell types. Therefore, traditional analysis of bulk gene-expression data might result in a loss of signal from less abundant cell types, reducing our ability to resolve unique features of cellular states and limiting the conclusions of an experiment. Traditional methods for studying cell heterogeneity, such as immunohistochemistry and flow cytometry, rely on a small number of phenotypic markers and are therefore limited in their utility for studying cancer heterogeneity.

The field of single-cell genomics has rapidly grown in recent years. Early methods for single-cell RNA sequencing (scRNA-seq) (e.g., SMART-seq, Fluidigm C1) relied on physically separating cells into microfluidics or high-density microtiter plates by flow cytometric cell sorting, which greatly limited the throughput. The emergence of methods based on high-throughput encapsulation of cells and barcoded capture beads into aqueous droplets within an oil emulsion (Macosko et al. 2015) has revolutionized the field, permitting whole transcriptome analysis of hundreds of thousands of cells. These methods generally use poly-A-based capture of mRNA and so are sensitive to RNA integrity. mRNA capture is followed by reverse transcription, fragmentation of cDNA, and short-read sequencing of the ends of fragments. To address the limitations of short-read sequencing, single-molecule sequencing on the PacBio or Oxford Nanopore systems can be used to study splicing and structural variants or to identify SNPs not contained within the sequenced fragments. For instance, single-molecule long-read sequencing of full-length cDNA libraries was used to analyze alternative splicing and somatic hypermutation events in T- and B-cell receptors in breast cancers (Singh et al. 2019). These methods can be scaled to whole transcriptomes, opening up the possibility of analyzing the association of splice variants, mRNA editing, or genomically encoded cancer variants at single-cell resolution.

scRNA-seq is challenging to implement on clinical specimens, as it typically requires intact, alive cells and thus is not compatible with flash-frozen or paraffin-embedded material. Although sequencing of nuclei from flash-frozen tissue can be applied to avoid dissociation methods (Habib et al. 2017) and better represent the composition of complex cellular mixtures (Slyper et al. 2020), these methods have several distinct drawbacks in their application to oncology. These drawbacks include that the data generated is typical of substantially reduced sensitivity compared to using whole cells, that samples cannot be processed to remove dead or low-quality cells, which are common in solid cancers, that subsets of interest cannot be prospectively selected, and that the approach is not compatible with cell-surface immunophenotyping methods such as CITE-seq (Stoeckius et al. 2017). The consequence is that most clinical studies have used freshly processed samples, which is logistically difficult, inefficient, and may exacerbate variability introduced by batch effects within the resulting data set. One approach to solve these challenges is to cryopreserve or fix samples, which allows later sample batching. Studies (Guillaumet-Adkins et al. 2017; Wu et al. 2021c) have shown that

cryopreservation effectively conserves the transcriptome and cell abundance seen in freshly processed breast cancer samples while maintaining maximum compatibility with downstream applications, including CITE-seq (Wu et al. 2021c), spatial transcriptomics (Wu et al. 2021b), and cell culture or xenotransplantation. The recent development of single-nuclei RNA sequencing (snRNA-seq) and scRNA-seq methods based on probe-based detection in fixed cells has markedly reduced the technical sensitivity to RNA integrity and will dramatically increase the application of these technologies to the analysis of clinical breast cancer cohorts (Vallejo et al. 2022).

It is increasingly apparent that cellular context plays a significant role in determining cellular phenotypes in breast cancer, such as heterogeneity in drug sensitivity or T-cell exhaustion. Understanding the organization of breast cancers requires complex high-dimensional methods to resolve the many cell states within a tumor and their expression of molecules. What was once an abstract dream has become a reality with the emergence of methods to perform spatially resolved transcriptomics. The ideal spatial transcriptomics platform would combine single-cell-level resolution, high-target multiplexing, high-sample throughput, and sensitivity with flexibility in sample preparation. Current technologies do not meet all these goals, often compromising on throughput, resolution, or sensitivity.

Until recently, commercially available platforms have used a variety of strategies to map transcripts to a spatial region, typically >50 μm in diameter and comprising numerous cells. These include the Visium platform from 10X Genomics and GeoMX DSP from Nanostring Technologies. Academic methods for spatial transcriptomics that use optical imaging of hybridized gene probes, such as Seq-FISH (Lubeck et al. 2014), have existed for approximately a decade and provide subcellular resolution for a panel of gene targets. However, these methods have generally been difficult for most laboratories to implement and often only work well on carefully prepared fresh tissues from model systems. However, numerous new commercial platforms converge on methods for spatial transcriptomics via probe hybridization that markedly improves the sensitivity and robustness of analyzing archival FFPE tissues. These platforms provide the capability to spatially resolve hundreds to thousands of genes and proteins at the subcellular resolution level in routine clinical specimens. The new capabilities offered by these platforms make it possible for the next generation of spatial methods to become the dominant cellular genomics technology, relegating scRNA-seq to specific applications.

A Model for Cellular Heterogeneity in Breast Cancer

The emergence of standardized methods for cellular genomics, particularly single-cell, and spatial transcriptomics, is revolutionizing our understanding of breast cancer heterogeneity. To understand breast cancer heterogeneity, it is helpful to take a view of breast cancers as cellular ecosystems, whereby the etiology of disease is a function of interactions among neoplastic, stromal, and immune cells together with extracellular molecules such as matrix proteins, cytokines, and growth factors. This problem may be viewed at three levels (Fig. 3). The first level is one of taxonomy, to define the cell types and states that occupy tumors, particularly those of clinical relevance, and determine molecular programs that define their behavior and which of them we may aim to manipulate for a therapeutic purpose. The second is to understand the communities or cellular hubs that cells form in tumors, the mechanisms driving heterotypic cellular coalescence, and the impact those interactions have on tumor phenotype. Finally, there are widespread differences in cellular content in breast cancers and recurrent combinations of cellular composition, or ecotypes, across tumors that have recently been reported (Luca et al. 2021; Wu et al. 2021b). However, the drivers and clinical significance of ecotypes are mostly unknown.

Breast Cancer Cellular Taxonomies

Single-cell and spatial transcriptomics have diverse applications in the study of breast cancer heterogeneity (Fig. 4). A major focus in this area has been on defining cell types and cellular states that comprise breast cancers using scRNA-

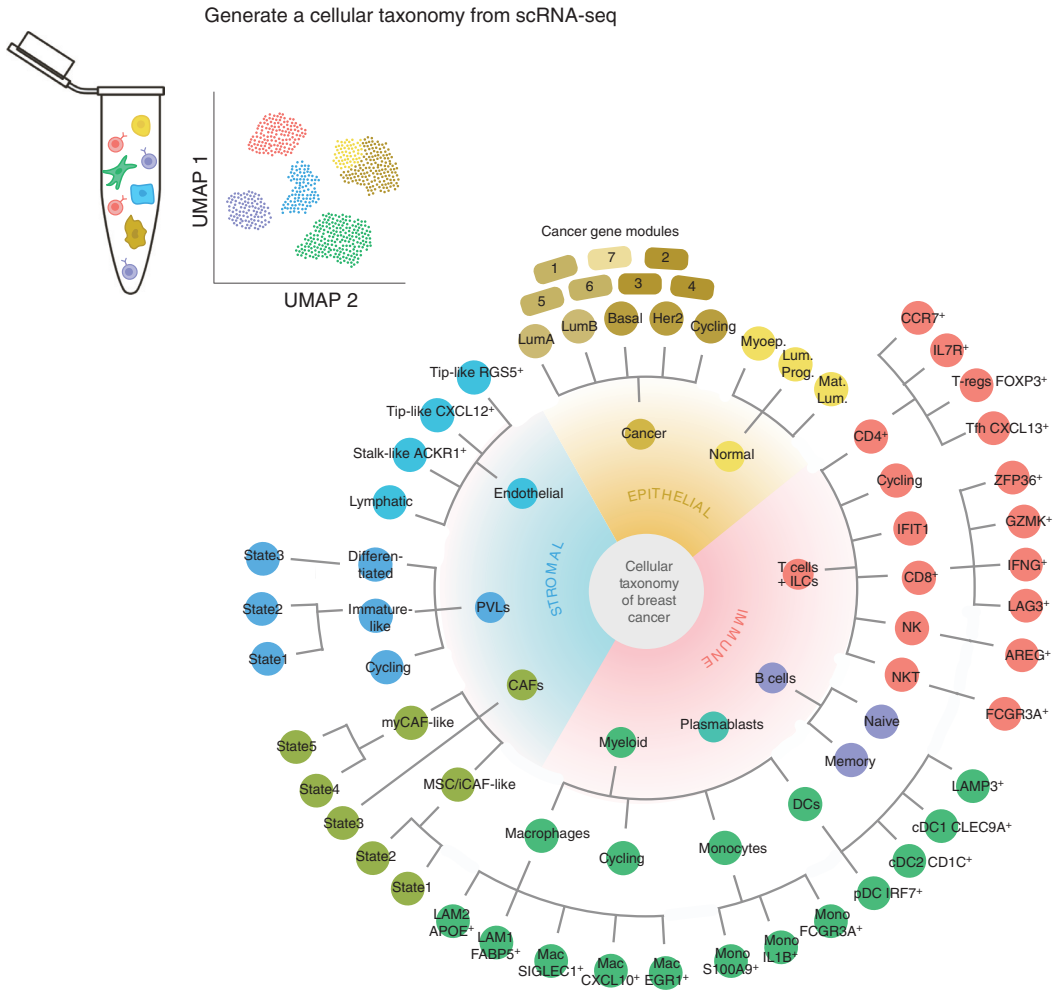

Figure 3. Breast cancer ecosystems can be viewed at the level of cellular taxonomies, cellular communities, and recurrent patterns of cellular composition, or ecotypes.

seq. The first generation of clinical scRNA-seq studies have typically produced unbiased descriptive atlases from early breast cancers (Chung et al. 2017; Singh et al. 2019; Qian et al. 2020; Wu et al. 2020, 2021; Gao et al. 2021; Pal et al. 2021), broadly sampling the cellular diversity of disease across neoplastic, immune, and stromal cells.

Neoplastic Cells

Heterogeneity and plasticity among neoplastic cells provide a source of diversity that drives drug resistance and are major challenges in the diagnosis and treatment of breast cancer. However, identifying the recurrent features of tumor heterogeneity by scRNA-seq is complicated by the dominance of patient-specific features when using unbiased clustering approaches (Pal et al. 2021; Wu et al. 2021b). The result is that cells from different patients typically cluster separately unless a stringent batch correction is applied, which can skew gene expression. Several approaches have been used to address this problem. The first is to apply the PAM50 intrinsic gene set to annotate individual tumor cell phenotypes. However, the majority of PAM50 genes return zero values at single-cell resolution in scRNA-seq data sets. An approach based on the concepts of intrinsic sub-

Cold Spring Harbor Perspectives in Medicine
www.perspectivesinmedicine.org

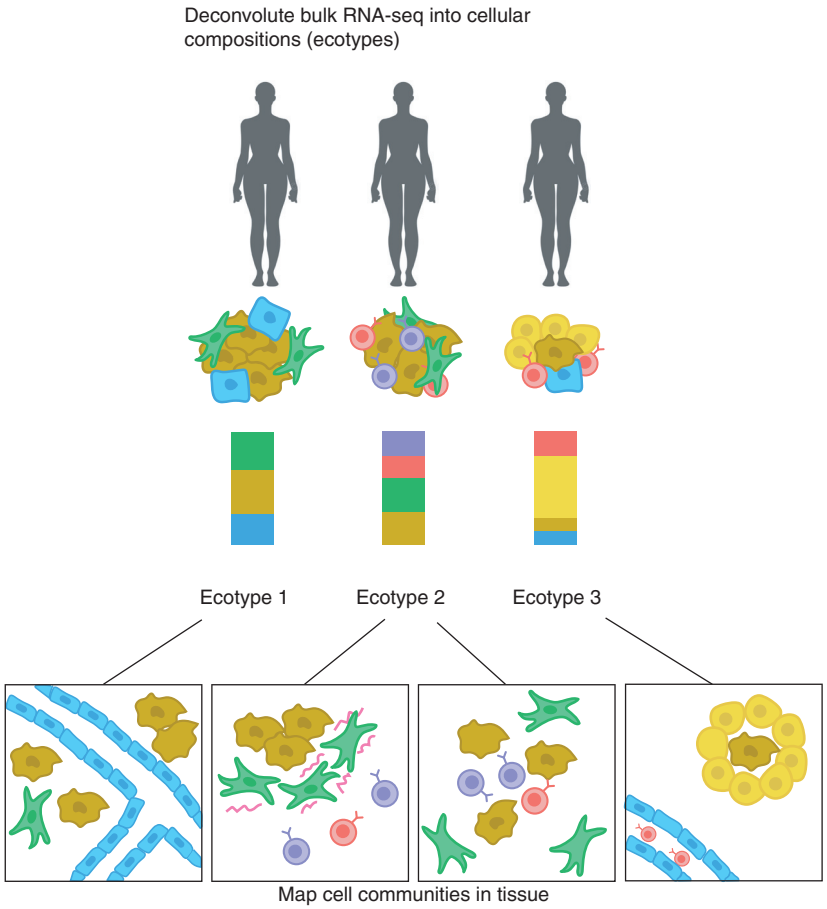

Figure 3. (*Continued.*)

types but trained on scRNA-seq data rather than bulk RNA-seq data, "scSubtyper," has been developed to address this issue, and the results reveal that intrinsic subtypes can be applied to individual cells. Notably, while the majority of cells within a tumor have subtypes concordant with the "bulk" annotation, a minor subset of cells with an scSubtype "discordant" to the bulk subtype is seen in a majority of cancers (Wu et al. 2021). Further work is needed to develop this method and test whether quantifying intratumoral heterogeneity with scSubtype predicts drug response or relapse (Wu et al. 2021b), where one would hypothesize that high intratumor heterogeneity subtype diversity might predict worse outcomes.

A second approach to define neoplastic heterogeneity is identifying modules of genes associated with intratumoral heterogeneity and then applying these modules across cancers (Wu et al. 2021b). Using modules, or gene sets, instead of individual genes was an idea developed in the bulk world that has been used widely, and it is likely this approach may be even more important in single-cell data where biological and technological noise is even greater. The result is a series of modules representing major biological features, such as proliferation, epithelial–mesenchymal transition (EMT), and ER signaling. Some of these clusters of gene modules are spatially segregated (Wu et al. 2021), suggesting either that the local microenvironment controls their expression or that they associate with epi/genomic subclones. Methods to infer DNA copy number states from scRNA-seq data have been developed (Patel et al. 2014; Gao et al. 2021) and will be important tools in linking genotype to

Cold Spring Harbor Perspectives in Medicine
www.perspectivesinmedicine.org

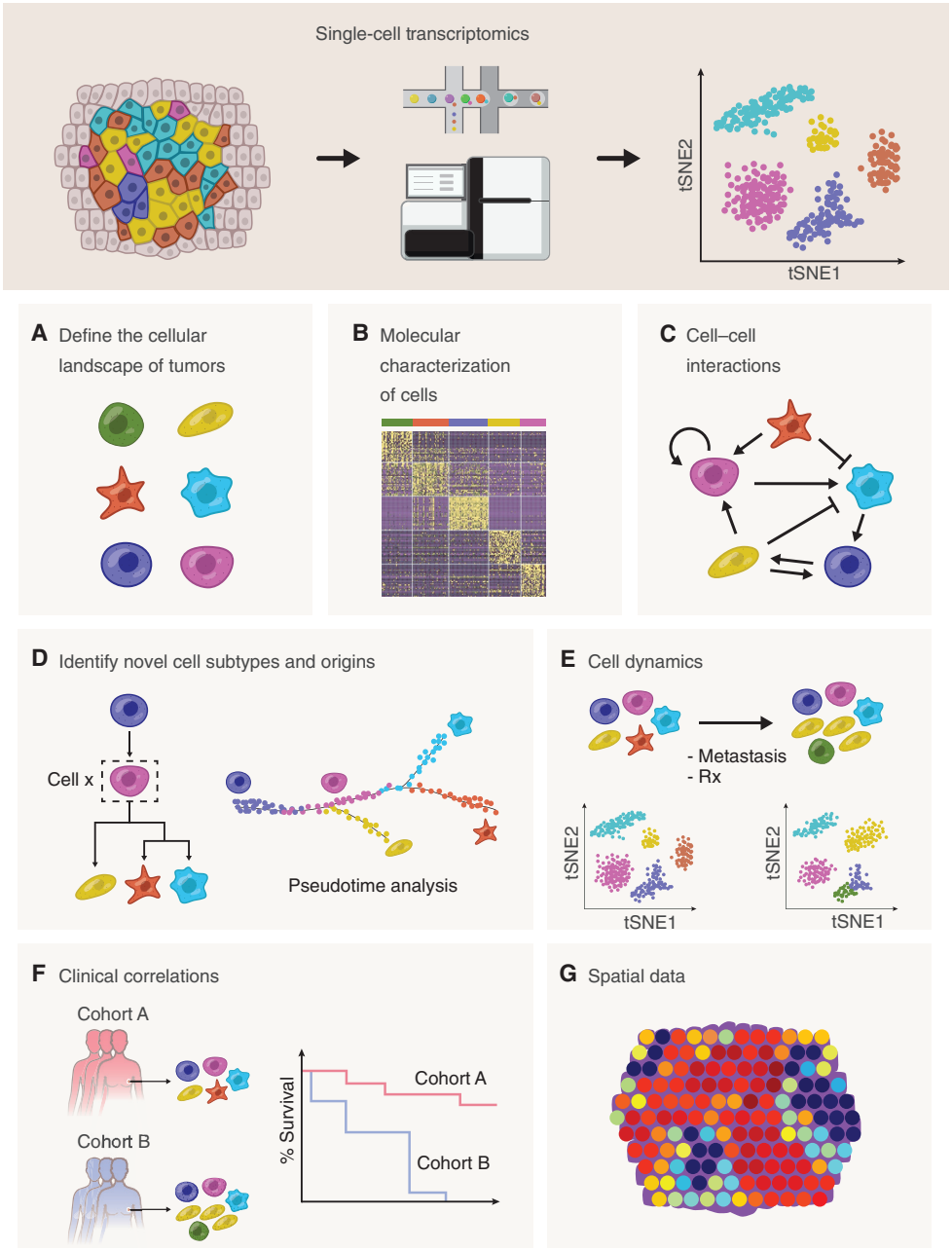

Figure 4. Single-cell transcriptomics can address a diversity of research questions. (*A*) Through integrated RNA and barcoded antibody measurements, cell types and states can be defined. (*B*) Gene-expression analysis can identify unique RNA and protein markers of populations. (*C*) Using algorithms that encompass known ligand–receptor pairings (Ramilowski et al. 2015; Efremova et al. 2020), putative cell–cell interactions can be identified. (*D*) Using dynamic changes in gene expression, branching trajectories of differentiation can be predicted (Trapnell et al. 2014). (*E*) By analyzing samples taken at different times or tissues, cellular dynamics during these processes can be discovered. (*F*) Cellular features of different clinical cohorts may reveal genes or signatures that are prognostic or predictive of response. (*G*) Patient- or disease-matched single-cell data allows accurate "deconvolution" of spatial data sampled from >1 cell into pseudo single-cell data. (tSNE) t-distributed stochastic neighbor embedding.

phenotype in single-cell and spatial transcriptomic data sets going forward.

Immune Cells

Many breast cancers contain an extensive immune infiltrate, and numerous immune cell populations are prognostic or predictive of response to chemotherapies and immunotherapies, typically T-cell and/or B-cell features (Hollern et al. 2019). Immune checkpoint inhibitors have made an impact on breast cancer, with these now being approved for use in metastatic and early-stage TNBC and being tested in all other breast clinical subtypes as well (see Bahl et al. 2022). Immune cell populations in breast cancer demonstrate substantially expanded phenotypic diversity compared to those in normal breast or blood (Azizi et al. 2018), suggesting that residence within malignant tissues provides an environment for diverse immune activation states. There are many similarities between the cell states of immune cells in breast cancer and those in other solid cancers, with the major variation being in relative abundance rather than unique cellular states. Analysis of resting and naive immune cell populations using scRNA-seq alone is challenging, as these cells are relatively transcriptionally inactive. It is here that multi-omic methods, such as CITE-seq (Stoeckius et al. 2017) that measure cell-surface epitopes using barcoded antibodies in parallel to transcriptomes, are particularly valuable in immunophenotyping (Al-Eryani et al. 2022).

T cells are the targets of currently approved immune checkpoint inhibitors. Several groups have reported a diverse repertoire of T cells along continuous spectra of activation, differentiation, and exhaustion within breast cancers (Azizi et al. 2018). Indeed, immune cell plasticity appears very high in breast cancer TME, resulting in novel populations not found outside of cancers (Al-Eryani et al. 2022), such as exhausted $CXCL13^{+}$ CD4 T follicular helper cells. T-cell states vary by breast cancer intrinsic subtype, with more proliferating and exhausted T cells found in TNBC (Pal et al. 2021; Wu et al. 2021b), mirroring the greater responses to immune checkpoint inhibitors in this clinical subgroup than other subgroups. Interestingly, individual expanded clones of T cells, identified by parallel V(D)J sequencing, occupy restricted spaces in this continuum, suggesting that signaling through the T-cell receptor is vital to defining gene-expression states. Tissue-resident memory CD8 T cells (Trm) are found in breast cancers, particularly ER-negative disease (Pal et al. 2021), and the signatures derived from CD103-positive CD8 T cells correlate with good prognosis, suggesting that Trm may mediate antitumor immune responses (Savas et al. 2018). However, caution must be used when interpreting the markers for Trm, which are not proven to mark tissue-resident cells in human cancer and may instead mark activation and/or exhaustion (Al-Eryani et al. 2022).

B cells and plasma cells are frequently present in breast cancers, particularly in HER2-positive and TNBC (Pal et al. 2021; Wu et al. 2021b), and B-cell abundance and immunoglobulin transcript levels are strongly associated with good outcomes (Iglesia et al. 2014; Hollern et al. 2019; Hu et al. 2021). The function of these cells in breast cancer is poorly known and subject to intense research activity. Most B cells in TNBC are memory B cells that have typically undergone affinity maturation (Singh et al. 2019; Hu et al. 2021). Clonally expanded B cells can be found in paired tumors and draining lymph nodes (Singh et al. 2019), suggesting antigen presentation, trafficking to tumors, and proliferation in response to antigens of these cells. The function of B cells in breast cancers and antigens to which they respond are not yet known and are important areas for future research.

Breast cancers contain a substantial number of myeloid cells across diverse states. Granulocytes such as neutrophils are poorly represented in scRNA-seq data sets, which may result from their low abundance, low transcript levels, and their sensitivity to the dissociation and capture protocols employed for scRNA-seq. Methods such as snRNA-seq are more likely to capture granulocytes. Macrophages in breast cancer form a number of clusters, including two substantial populations of TREM2-positive cells

that resemble lipid-associated macrophages initially identified in obese mice and humans (Jaitin et al. 2019). These populations, now found in many types of solid cancers (Mulder et al. 2021), are notable for their expression of lipid metabolism molecules and immunosuppressive molecules, such as CCL18, PDL1, and PDL2, and their association with poor patient survival (Wu et al. 2021b). TREM2-positive macrophages may be essential to understanding breast cancer immunity, as their abundance is associated with response to PDL1 blockade in TNBC (Zheng et al. 2021) and their depletion can sensitize mouse models of breast cancer to immune checkpoint inhibition (Molgora et al. 2020).

Stromal Cells

Breast cancers exhibit a diverse stromal milieu, and breast cancer specimens and models have been at the forefront of research into the role of stroma in cancer for decades. More recently, scRNA-seq has revealed surprising new insights into stromal biology. It is now clear that cancer-associated fibroblasts (CAFs) adopt diverse phenotypes, ranging from myofibroblasts with features of contractility and extracellular matrix remodeling through to less-differentiated cells resembling mesenchymal stromal cells. Trajectory analysis (Fig. 4D) suggests a continuum of differentiation states between these two extreme states (Wu et al. 2021b). These two states are commonly called myofibroblastic CAF (myCAF) and inflammatory CAF (iCAF), adopting the nomenclatures for similar cells reported in pancreatic cancer (Sahai et al. 2020) with which they share many similarities. iCAFs express many immunoregulatory molecules, including immune checkpoint ligands PDL1 and PDL2 and chemokines such as CXCL12, and colocalize with T and B lymphocytes, supporting the suggestion that they may be essential players in immunoregulation within tumors. Although iCAF may be the progenitor of myCAF, they occupy different regions of breast cancers, with myCAF typically closely associated with cancer cells (Wu et al. 2021b). The distinct localization of these cells suggests a role for unique microenvironments in either differentially recruiting CAF subsets or regulating CAF differentiation.

scRNA-seq has also revealed diversities within other stromal cell populations in the TME. Endothelial cells lie across a spectrum of differentiation, from cells resembling stalk-like cells to tip-like cells and a subset characterized by migration and angiogenic signatures (Wu et al. 2021b). Cells with features of smooth muscle cells and pericytes, termed perivascular-like cells (PVLs), are as frequent as CAFs in early breast cancers (Wu et al. 2021a). PVLs are also found across a range of differentiation states. Surprisingly, PVLs can be frequently found distant from blood vessels in breast cancers (Wu et al. 2021a). Given the overlapping morphology and markers with CAFs (e.g., smooth muscle actin), PVLs have almost certainly been misclassified as CAFs in the past. The function of PVL, particularly those disseminated throughout the tumor parenchyma, is a major knowledge gap.

A recurring observation among these studies is that rather than occupying discrete states, or cell types, cells of all three major lineages differ across a continuous spectrum of phenotypes or states, representing multiple axes comprising differentiation, activation, and others. Cells are presumably transitioning dynamically between states, so a significant challenge will be to understand the drivers of this plasticity, be it intrinsic such as genomic, or extrinsic such as extracellular interactions. These insights will drive a new generation of therapeutics based on manipulating the cellular composition or signaling within the TME.

Cellular Communities

While the field has long recognized the existence of breast cancer cell heterogeneity, few studies have illuminated its topological associations. A recent pan-cancer analysis examining associations between cancer gene-expression modules and cellular contexts revealed the regulation of interferon response signatures in cancer cells by adjacent macrophages and T cells (Barkley et al. 2022), which was also observed in HER2-positive breast cancers (Andersson et al. 2021). Recently, iCAFs were shown to associate with T and B lymphocytes, and spatial-resolved ligand-receptor

mapping revealed potential signaling intermediates between these populations (Wu et al. 2020).

Tertiary lymphoid structures (TLS) are complex cellular structures comprised of lymphocytes, stromal cells, and myeloid cells that resemble secondary lymphoid organs such as lymph nodes, and they are observed in some breast tumors. Most cancers with TLS are associated with superior prognosis and response to immunotherapy (Schumacher and Thommen 2022). It is likely that many important stromal and immune cell types identified recently through single-cell methods, including memory B cells, iCAFs, CD4 T follicular helper cells, and dendritic cell subsets, interact within TLS to alter the antitumor immunity. The likely importance of these structures in the breast cancer immune context demonstrates the potential for spatial transcriptomics in mapping breast cancer ecosystems (Andersson et al. 2021). By placing cells within an actionable distance, spatial transcriptomics may provide an opportunity for new drug target discovery.

Ecotypes

Single-cell transcriptomics has revealed highly variable cell-type compositions among breast cancers (Pal et al. 2021; Wu et al. 2021b), which raises the question of whether these variabilities in cell-type composition are the result of a generally random process or instead reflects recurring patterns of cellular frequencies. Because single-cell atlases to date contain no more than 34 cases of breast cancer (Pal et al. 2021), it is not possible to address this question robustly yet. However, single-cell reference profiles of every cell type can be used to infer the proportion of cell types in large, published bulk RNA-seq data sets (Fig. 5).

Several computational methods have been developed to infer the abundance of different

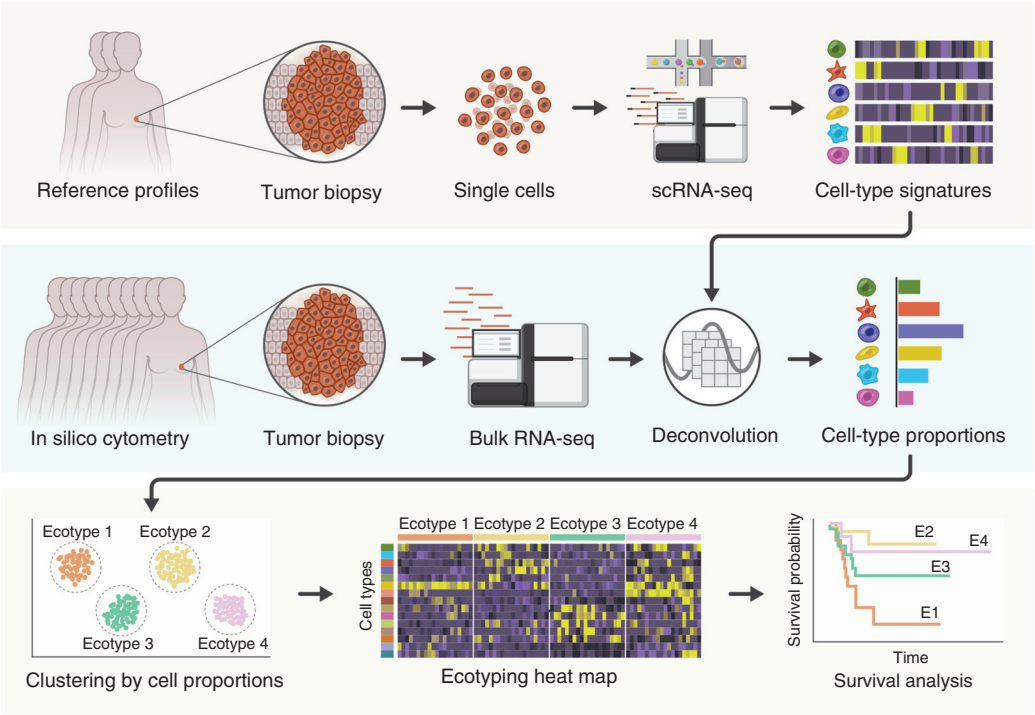

Figure 5. Disease-specific single-cell transcriptomics references can be used to predict cellular frequencies within bulk RNA sequencing (RNA-seq) data sets via in silico cytometry. This strategy can be used to study cellular proportions across large clinical cohorts.

cell types in bulk tissue. One example is CIBERSORT, a deconvolution method based on support vector regression that allows the characterization of the immune cell composition of bulk tissues (Newman et al. 2015). CIBERSORT has been demonstrated to have prognostic value in breast cancer (Craven et al. 2021). Methods such as CIBERSORT use canonical gene signatures, often derived from blood, for deconvolution. Because of the marked influence of the TME on cellular phenotypes, the accuracy of predictions is markedly increased by using high-resolution and disease-specific reference profiles. Using scRNA-seq data from 26 breast cancer cases as reference, ~2000 cases of breast cancer in the Metabric cohort (Curtis et al. 2012) have been deconvoluted to reveal a number of recurrent compositional clusters, which are called ecotypes (Wu et al. 2021b). Ecotypes are composed of immune, stromal, and neoplastic cells in recurring proportions, suggesting that all these three cell lineages contribute to building these recurrent tumor neighborhoods. Ecotypes, in part, recapitulate the PAM50 and IntClusters derived from the Metabric and build upon these by integrating the TME and in finding additional biological groups that are associated with prognosis. Ecotypes appear to be conserved across diverse epithelial tumor types (Luca et al. 2021), but the mechanisms organizing ecotypes are mostly unknown. Although it remains to be tested, ecotypes may predict response to therapy, particularly immunotherapy.

FUTURE DIRECTIONS

In many ways, the age of cellular- and spatial-resolved "omics" is only beginning. A challenge in the single-cell space is to systematically define cell states and a common ontology while avoiding duplication of nomenclature. This will require integrative studies that find the shared features across studies and disease subsets and rely on methods that enable less biased capture of cell types and states, particularly for complex cell types such as granulocytes, neurons, and adipocytes, while retaining sensitivity. Methods compatible with fixed cells, particularly those embedded in paraffin, will open up large clinically valuable cohorts to cellular genomics to definitively test the value of these new technologies to translational oncology.

Current methods of single-cell analysis typically sample ~1000–10,000 cells per tumor, and the majority of studies have captured several cell types, so any one cell type is shallowly sampled. Future work in this area will need to capture far more cells per tumor or focus on specific cellular populations to more comprehensively resolve cellular states and plasticity. In breast cancer, this is particularly relevant for understanding neoplastic cell heterogeneity. For instance, ER, PR, and androgen receptor (AR) are critically important to breast cancer biology and treatment. Intriguingly, histochemical studies reveal that ER, PR, and AR are generally heterogeneously expressed in breast cancers (Lindström et al. 2018), yet cancers with only 1% ER-positive cells may respond to endocrine therapies. Single-cell transcriptomics, in particular, when combined with newer methods such as single-cell ATAC-seq, may provide a path to better understand endocrine receptor function and endocrine therapy responsiveness.

Cellular genomics has the potential to unveil insights into the cellular basis for drug response and resistance. Although most studies have focused on treatment-naive disease, Kim and colleagues used scRNA-seq to monitor the cellular dynamics of TNBC in patients undergoing neoadjuvant chemotherapy treatment (Kim et al. 2018). Although this study cannot resolve the dynamics of stromal and immune cells through therapy, it reveals dynamic changes in neoplastic cell gene expression in samples from patients who had an incomplete response, including activation of pathways relating to extracellular matrix remodeling, AKT signaling, and EMT. Similarly, Griffiths et al. (2021) applied scRNA-seq to ER-positive breast cancers from a cohort of patients receiving the endocrine therapy letrozole with or without the CDK4/6 inhibitor ribociclib. This analysis revealed the emergence of conserved resistance mechanisms, including ER activation downstream of ERBB4 or JNK activation downstream of FGFR2 activation. This dynamic plasticity observed across cancer types and drug classes is a significant challenge for cancer thera-

py and is an area of investigation where cellular genomics will be a powerful tool. These insights may inform future trials of rational combination therapy and establish the groundwork to formally evaluate the benefit of cellular genomics to translational science through the analysis of substantial clinical trial cohorts.

REFERENCES

*Reference is also in this subject collection.

Albain KS, Barlow WE, Shak S, Hortobagyi GN, Livingston RB, Yeh IT, Ravdin P, Bugarini R, Baehner FL, Davidson NE, et al. 2010. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol* **11:** 55–65. doi:10.1016/S1470-2045(09)70314-6

Al-Eryani G, Bartonicek N, Chan CL, Anderson A, Harvey K, Wu SZ, Roden D, Wang T, Reeves J, Yeung BZ, et al. 2022. Integration of single-cell RNA and protein data identifies novel clinically-relevant lymphocyte phenotypes in breast cancers. bioRxiv doi:10.1101/2022.05.31.494081

Andersson A, Larsson L, Stenbeck L, Salmén F, Ehinger A, Wu SZ, Al-Eryani G, Roden D, Swarbrick A, Borg A, et al. 2021. Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. *Nat Commun* **12:** 6012. doi:10.1038/s41467-021-26271-2

Andre F, Ismaila N, Allison KH, Barlow WE, Collyar DE, Damodaran S, Henry NL, Jhaveri K, Kalinsky K, Kuderer NM, et al. 2022. Biomarkers for adjuvant endocrine and chemotherapy in early-stage breast cancer: ASCO guideline update. *J Clin Oncol* **40:** 1816–1837. doi:10.1200/JCO.22.00069

Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, Nainys J, Wu K, Kiseliovas V, Setty M, et al. 2018. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* **174:** 1293–1308.e36. doi:10.1016/j.cell.2018.05.060

* Bahl S, Carroll JS, Lupien M. 2022. Chromatin variants reveal the genetic determinants of oncogenesis in breast cancer. *Cold Spring Harb Perspect Med* **12:** a041322. doi:10.1101/cshperspect.a041322

Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, Lawrence MS, Sivachenko AY, Sougnez C, Zou L, et al. 2012. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* **486:** 405–409. doi:10.1038/nature11154

Barkley D, Moncada R, Pour M, Liberman DA, Dryg I, Werba G, Wang W, Baron M, Rao A, Xia B, et al. 2022. Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment. *Nat Genet* **54:** 1192–1201. doi:10.1038/s41588-022-01141-9

Bartlett J, Bayani J, Marshall A, Dunn JA, Campbell A, Cunningham C, Sobol MS, Hall PS, Poole CJ, Cameron DA, et al. 2016. Comparing breast cancer multiparameter tests in the OPTIMA prelim trial: no test is more equal than the others. *J Natl Cancer Inst* **108:** djw050. doi:10.1093/jnci/djw050

Buus R, Sestak I, Kronenwett R, Denkert C, Dubsky P, Krappmann K, Scheer M, Petry C, Cuzick J, Dowsett M. 2016. Comparison of EndoPredict and EPclin with Oncotype DX recurrence score for prediction of risk of distant recurrence after endocrine therapy. *J Natl Cancer Inst* **108:** djw149. doi:10.1093/jnci/djw149

Buyse M, Loi S, Van't Veer L, Viale G, Delorenzi M, Glas AM, Saghatchian d'Assignies M, Bergh J, Lidereau R, Ellis P, et al. 2006. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* **98:** 1183–1192. doi:10.1093/jnci/djj329

Cancer Genome Atlas Network. 2012. Comprehensive molecular portraits of human breast tumours. *Nature* **490:** 61–70. doi:10.1038/nature11412

Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, Pierga JY, Brain E, Causeret S, DeLorenzi M, et al. 2016. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med* **375:** 717–729. doi:10.1056/NEJMoa1602253

Cardoso F, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rubio I, Zackrisson S, Senkus E. 2019. Early breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol* **30:** 1194–1220. doi:10.1093/annonc/mdz173

Chung W, Eum HH, Lee HO, Lee KM, Lee HB, Kim KT, Ryu HS, Kim S, Lee JE, Park YH, et al. 2017. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat Commun* **8:** 1–12. doi:10.1038/ncomms15081

Ciriello G, Sinha R, Hoadley KA, Jacobsen AS, Reva B, Perou CM, Sander C, Schultz N. 2013. The molecular diversity of luminal A breast tumors. *Breast Cancer Res Treat* **141:** 409–420. doi:10.1007/s10549-013-2699-3

Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, Zhang H, McLellan M, Yau C, Kandoth C, et al. 2015. Comprehensive molecular portraits of invasive lobular breast cancer. *Cell* **163:** 506–519. doi:10.1016/j.cell.2015.09.033

Craven KE, Gökmen-Polar Y, Badve SS. 2021. CIBERSORT analysis of TCGA and METABRIC identifies subgroups with better outcomes in triple negative breast cancer. *Sci Rep* **11:** 1–19. doi:10.1038/s41598-021-83913-7

Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, et al. 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486:** 346–352. doi:10.1038/nature10983

Dowsett M, Cuzick J, Wale C, Forbes J, Mallon EA, Salter J, Quinn E, Dunbier A, Baum M, Buzdar A, et al. 2010. Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: a TransATAC study. *J Clin Oncol* **28:** 1829–1834. doi:10.1200/JCO.2009.24.4798

Drukker CA, Bueno-de-Mesquita J, Retèl VP, van Harten WH, van Tinteren H, Wesseling J, Roumen RM, Knauer M, van't Veer LJ, Sonke GS, et al. 2013. A prospective evaluation of a breast cancer prognosis signature in the

Cold Spring Harbor Perspectives in Medicine
www.perspectivesinmedicine.org

observational RASTER study. *Int J Cancer* **133:** 929–936. doi:10.1002/ijc.28082

Dubsky P, Brase J, Jakesz R, Rudas M, Singer C, Greil R, Dietze O, Luisser I, Klug E, Sedivy R, et al. 2013. The EndoPredict score provides prognostic information on late distant metastases in $ER^+/HER2^-$ breast cancer patients. *Br J Cancer* **109:** 2959–2964. doi:10.1038/bjc.2013.671

Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. 2020. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat Protoc* **15:** 1484–1506. doi:10.1038/s41596-020-0292-x

Fan C, Prat A, Parker JS, Liu Y, Carey LA, Troester MA, Perou CM. 2011. Building prognostic models for breast cancer patients using clinical variables and hundreds of gene expression signatures. *BMC Med Genomics* **4:** 3. doi:10.1186/1755-8794-4-3

Gao R, Bai S, Henderson YC, Lin Y, Schalck A, Yan Y, Kumar T, Hu M, Sei E, Davis A, et al. 2021. Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat Biotechnol* **39:** 599–608. doi:10.1038/s41587-020-00795-2

Garcia-Recio S, Thennavan A, East MP, Parker JS, Cejalvo JM, Garay JP, Hollern DP, He X, Mott KR, Galván P, et al. 2020. FGFR4 regulates tumor subtype differentiation in luminal breast cancer and metastatic disease. *J Clin Invest* **130:** 4871–4887. doi:10.1172/JCI130323

Gatza ML, Lucas JE, Barry WT, Kim JW, Wang Q, Crawford MD, Datto MB, Kelley M, Mathey-Prevot B, Potti A, et al. 2010. A pathway-based classification of human breast cancer. *Proc Natl Acad Sci* **107:** 6994–6999. doi:10.1073/pnas.0912708107

Gnant M, Filipits M, Greil R, Stoeger H, Rudas M, Bago-Horvath Z, Mlineritsch B, Kwasny W, Knauer M, Singer C, et al. 2014. Predicting distant recurrence in receptor-positive breast cancer patients with limited clinicopathological risk: using the PAM50 risk of recurrence score in 1478 postmenopausal patients of the ABCSG-8 trial treated with adjuvant endocrine therapy alone. *Ann Oncol* **25:** 339–345. doi:10.1093/annonc/mdt494

Griffiths JI, Chen J, Cosgrove PA, O'Dea A, Sharma P, Ma C, Trivedi M, Kalinsky K, Wisinski KB, O'Regan R, et al. 2021. Serial single-cell genomics reveals convergent subclonal evolution of resistance as patients with early-stage breast cancer progress on endocrine plus CDK4/6 therapy. *Nat Cancer* **2:** 658–671. doi:10.1038/s43018-021-00215-7

Guillaumet-Adkins A, Rodríguez-Esteban G, Mereu E, Mendez-Lago M, Jaitin DA, Villanueva A, Vidal A, Martinez-Marti A, Felip E, Vivancos A, et al. 2017. Single-cell transcriptome conservation in cryopreserved cells and tissues. *Genome Biol* **18:** 1–15. doi:10.1186/s13059-017-1171-9

Habib N, Avraham-Davidi I, Basu A, Burks T, Shekhar K, Hofree M, Choudhury SR, Aguet F, Gelfand E, Ardlie K, et al. 2017. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods* **14:** 955–958. doi:10.1038/nmeth.4407

Heng YJ, Lester SC, Tse GM, Factor RE, Allison KH, Collins LC, Chen YY, Jensen KC, Johnson NB, Jeong JC, et al. 2017. The molecular basis of breast cancer pathological phenotypes. *J Pathol* **241:** 375–391. doi:10.1002/path.4847

Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, Shen R, Taylor AM, Cherniack AD, Thorsson V, et al. 2018. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* **173:** 291–304.e6. doi:10.1016/j.cell.2018.03.022

Hollern DP, Xu N, Thennavan A, Glodowski C, Garcia-Recio S, Mott KR, He X, Garay JP, Carey-Ewend K, Marron D, et al. 2019. B cells and T follicular helper cells mediate response to checkpoint inhibitors in high mutation burden mouse models of breast cancer. *Cell* **179:** 1191–1206.e21. doi:10.1016/j.cell.2019.10.028

Hu Q, Hong Y, Qi P, Lu G, Mai X, Xu S, He X, Guo Y, Gao L, Jing Z, et al. 2021. Atlas of breast cancer infiltrated B-lymphocytes revealed by paired single-cell RNA-sequencing and antigen receptor profiling. *Nat Commun* **12:** 1–13. doi:10.1038/s41467-021-22300-2

Huo D, Hu H, Rhie SK, Gamazon ER, Cherniack AD, Liu J, Yoshimatsu TF, Pitt JJ, Hoadley KA, Troester M, et al. 2017. Comparison of breast cancer molecular features and survival by African and European ancestry in The Cancer Genome Atlas. *JAMA Oncol* **3:** 1654–1662. doi:10.1001/jamaoncol.2017.0595

ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. 2020. Pan-cancer analysis of whole genomes. *Nature* **578:** 82–93. doi:10.1038/s41586-020-1969-6

Iglesia MD, Vincent BG, Parker JS, Hoadley KA, Carey LA, Perou CM, Serody JS. 2014. Prognostic B-cell signatures using mRNA-seq in patients with subtype-specific breast and ovarian cancer. *Clin Cancer Res* **20:** 3818–3829. doi:10.1158/1078-0432.CCR-13-3368

Jaitin DA, Adlung L, Thaiss CA, Weiner A, Li B, Descamps H, Lundgren P, Bleriot C, Liu Z, Deczkowska A, et al. 2019. Lipid-associated macrophages control metabolic homeostasis in a Trem2-dependent manner. *Cell* **178:** 686–698.e14. doi:10.1016/j.cell.2019.05.054

Kalinsky K, Barlow WE, Gralow JR, Meric-Bernstam F, Albain KS, Hayes DF, Lin NU, Perez EA, Goldstein LJ, Chia SK, et al. 2021. 21-gene assay to inform chemotherapy benefit in node-positive breast cancer. *N Engl J Med* **385:** 2336–2347. doi:10.1056/NEJMoa2108873

Kim C, Gao R, Sei E, Brandt R, Hartman J, Hatschek T, Crosetto N, Foukakis T, Navin NE. 2018. Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing. *Cell* **173:** 879–893.e13. doi:10.1016/j.cell.2018.03.041

Krug K, Jaehnig EJ, Satpathy S, Blumenberg L, Karpova A, Anurag M, Miles G, Mertins P, Geffen Y, Tang LC, et al. 2020. Proteogenomic landscape of breast cancer tumorigenesis and targeted therapy. *Cell* **183:** 1436–1456.e31. doi:10.1016/j.cell.2020.10.036

Lindström LS, Yau C, Czene K, Thompson CK, Hoadley KA, Van't Veer LJ, Balassanian R, Bishop JW, Carpenter PM, Chen YY, et al. 2018. Intratumor heterogeneity of the estrogen receptor and the long-term risk of fatal breast cancer. *J Natl Cancer Inst* **110:** 726–733. doi:10.1093/jnci/djx270

Lubeck E, Coskun AF, Zhiyentayev T, Ahmad M, Cai L. 2014. Single-cell in situ RNA profiling by sequential hy-

Cold Spring Harbor Perspectives in Medicine
www.perspectivesinmedicine.org

bridization. *Nat Methods* **11:** 360–361. doi:10.1038/nmeth.2892

Luca BA, Steen CB, Matusiak M, Azizi A, Varma S, Zhu C, Przybyl J, Espín-Pérez A, Diehn M, Alizadeh AA, et al. 2021. Atlas of clinically distinct cell states and ecosystems across human solid tumors. *Cell* **184:** 5482–5496.e28. doi:10.1016/j.cell.2021.09.014

Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al. 2015. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161:** 1202–1214. doi:10.1016/j.cell.2015.05.002

Massagué J. 2007. Sorting out breast-cancer gene signatures. *N Engl J Med* **356:** 294–297. doi:10.1056/NEJMe068292

Molgora M, Esaulova E, Vermi W, Hou J, Chen Y, Luo J, Brioschi S, Bugatti M, Omodei AS, Ricci B, et al. 2020. TREM2 modulation remodels the tumor myeloid landscape enhancing anti-PD-1 immunotherapy. *Cell* **182:** 886–900.e17. doi:10.1016/j.cell.2020.07.013

Mulder K, Patel AA, Kong WT, Piot C, Halitzki E, Dunsmore G, Khalilnezhad S, Irac SE, Dubuisson A, Chevrier M, et al. 2021. Cross-tissue single-cell landscape of human monocytes and macrophages in health and disease. *Immunity* **54:** 1883–1900.e5. doi:10.1016/j.immuni.2021.07.007

Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. 2015. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12:** 453–457. doi:10.1038/nmeth.3337

Okoniewski MJ, Miller CJ. 2006. Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. *BMC Bioinformatics* **7:** 1–14. doi:10.1186/1471-2105-7-276

Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, et al. 2004. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* **351:** 2817–2826. doi:10.1056/NEJMoa041588

Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, Cronin M, Baehner FL, Watson D, Bryant J, et al. 2006. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor–positive breast cancer. *J Clin Oncol* **24:** 3726–3734. doi:10.1200/JCO.2005.04.7985

Pal B, Chen Y, Vaillant F, Capaldo BD, Joyce R, Song X, Bryant VL, Penington JS, Di Stefano L, Tubau Ribera N, et al. 2021. A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. *EMBO J* **40:** e107333. doi:10.15252/embj.2020107333

Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, Cahill DP, Nahed BV, Curry WT, Martuza RL, et al. 2014. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344:** 1396–1401. doi:10.1126/science.1254257

Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, et al. 2000. Molecular portraits of human breast tumours. *Nature* **406:** 747–752. doi:10.1038/35021093

Qian J, Olbrecht S, Boeckx B, Vos H, Laoui D, Etlioglu E, Wauters E, Pomella V, Verbandt S, Busschaert P, et al. 2020. A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. *Cell Res* **30:** 745–762. doi:10.1038/s41422-020-0355-0

Ramilowski JA, Goldberg T, Harshbarger J, Kloppmann E, Lizio M, Satagopam VP, Itoh M, Kawaji H, Carninci P, Rost B, et al. 2015. A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat Commun* **6:** 1–12. doi:10.1038/ncomms8866

Saal LH, Vallon-Christersson J, Häkkinen J, Hegardt C, Grabau D, Winter C, Brueffer C, Tang MH, Reuterswärd C, Schulz R, et al. 2015. The Sweden Cancerome Analysis Network—Breast (SCAN-B) Initiative: a large-scale multicenter infrastructure towards implementation of breast cancer genomic analyses in the clinical routine. *Genome Med* **7:** 20. doi:10.1186/s13073-015-0131-9

Sahai E, Astsaturov I, Cukierman E, DeNardo DG, Egeblad M, Evans RM, Fearon D, Greten FR, Hingorani SR, Hunter T, et al. 2020. A framework for advancing our understanding of cancer-associated fibroblasts. *Nat Rev Cancer* **20:** 174–186. doi:10.1038/s41568-019-0238-1

Savas P, Virassamy B, Ye C, Salim A, Mintoff CP, Caramia F, Salgado R, Byrne DJ, Teo ZL, Dushyanthen S, et al. 2018. Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. *Nat Med* **24:** 986 993. doi:10.1038/s41591-018-0078-7

Schumacher TN, Thommen DS. 2022. Tertiary lymphoid structures in cancer. *Science* **375:** eabf9419. doi:10.1126/science.abf9419

Sgroi DC, Carney E, Zarrella E, Steffel L, Binns SN, Finkelstein DM, Szymonifka J, Bhan AK, Shepherd LE, Zhang Y, et al. 2013a. Prediction of late disease recurrence and extended adjuvant letrozole benefit by the HOXB13/IL17BR biomarker. *J Natl Cancer Inst* **105:** 1036–1042. doi:10.1093/jnci/djt146

Sgroi DC, Sestak I, Cuzick J, Zhang Y, Schnabel CA, Schroeder B, Erlander MG, Dunbier A, Sidhu K, Lopez-Knowles E, et al. 2013b. Prediction of late distant recurrence in patients with oestrogen-receptor-positive breast cancer: a prospective comparison of the breast-cancer index (BCI) assay, 21-gene recurrence score, and IHC4 in the TransATAC study population. *Lancet Oncol* **14:** 1067–1076. doi:10.1016/S1470-2045(13)70387-5

Singh M, Al-Eryani G, Carswell S, Ferguson JM, Blackburn J, Barton K, Roden D, Luciani F, Giang Phan T, Junankar S, et al. 2019. High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. *Nat Commun* **10:** 1–13. doi:10.1038/s41467-019-11049-4

Slyper M, Porter C, Ashenberg O, Waldman J, Drokhlyansky E, Wakiro I, Smillie C, Smith-Rosario G, Wu J, Dionne D, et al. 2020. A single-cell and single-nucleus RNA-seq toolbox for fresh and frozen human tumors. *Nat Med* **26:** 792–802. doi:10.1038/s41591-020-0844-1

Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, Van De Rijn M, Jeffrey SS, et al. 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci* **98:** 10869–10874. doi:10.1073/pnas.191367098

Sotiriou C, Pusztai L. 2009. Gene-expression signatures in breast cancer. *N Engl J Med* **360:** 790–800. doi:10.1056/NEJMra0801289

Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, Geyer CE Jr, Dees EC, Goetz MP, Olson JA Jr,

et al. 2018. Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. *N Engl J Med* **379:** 111–121. doi:10.1056/NEJMoa1804710

Staaf J, Häkkinen J, Hegardt C, Saal LH, Kimbung S, Hedenfalk I, Lien T, Sørlie T, Naume B, Russnes H, et al. 2022. RNA sequencing-based single sample predictors of molecular subtype and risk of recurrence for clinical assessment of early-stage breast cancer. *NPJ Breast Cancer* **8:** 94. doi:10.1038/s41523-022-00465-3

Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, Satija R, Smibert P. 2017. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* **14:** 865–868. doi:10.1038/nmeth.4380

Tanioka M, Fan C, Parker JS, Hoadley KA, Hu Z, Li Y, Hyslop TM, Pitcher BN, Soloway MG, Spears PA, et al. 2018. Integrated analysis of RNA and DNA from the phase III trial CALGB 40601 identifies predictors of response to trastuzumab-based neoadjuvant chemotherapy in HER2-positive breast cancer. *Clin Cancer Res* **24:** 5292–5304. doi:10.1158/1078-0432.CCR-17-3431

Thennavan A, Beca F, Xia Y, Recio SG, Allison K, Collins LC, Tse GM, Chen YY, Schnitt SJ, Hoadley KA, et al. 2021. Molecular analysis of TCGA breast cancer histologic types. *Cell Genom* **1:** 100067. doi:10.1016/j.xgen.2021.100067

Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32:** 381–386. doi:10.1038/nbt.2859

Vallejo AF, Harvey K, Wang T, Wise K, Butler LM, Polo J, Plummer J, Swarbrick A, Martelotto LG. 2022. snPATHO-seq: unlocking the FFPE archives for single nucleus RNA profiling. bioRxiv doi:10.1101/2022.08.23.505054

Van't Veer LJ, Dai H, Van De Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, Van Der Kooy K, Marton MJ, Witteveen AT, et al. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415:** 530–536. doi:10.1038/415530a

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10:** 57–63. doi:10.1038/nrg2484

Wu SZ, Roden DL, Wang C, Holliday H, Harvey K, Cazet AS, Murphy KJ, Pereira B, Al-Eryani G, Bartonicek N, et al. 2020. Stromal cell diversity associated with immune evasion in human triple-negative breast cancer. *EMBO J* **39:** e104063.

Wu F, Yang J, Liu J, Wang Y, Mu J, Zeng Q, Deng S, Zhou H. 2021a. Signaling pathways in cancer-associated fibroblasts and targeted therapy for cancer. *Signal Transduct Target Ther* **6:** 218. doi:10.1038/s41392-021-00641-0

Wu SZ, Al-Eryani G, Roden DL, Junankar S, Harvey K, Andersson A, Thennavan A, Wang C, Torpy JR, Bartonicek N, et al. 2021b. A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet* **53:** 1334–1347. doi:10.1038/s41588-021-00911-1

Wu SZ, Roden DL, Al-Eryani G, Bartonicek N, Harvey K, Cazet AS, Chan CL, Junankar S, Hui MN, Millar EA, et al. 2021c. Cryopreservation of human cancers conserves tumour heterogeneity for single-cell multi-omics analysis. *Genome Med* **13:** 1–17. doi:10.1186/s13073-021-00885-z

Zheng L, Qin S, Si W, Wang A, Xing B, Gao R, Ren X, Wang L, Wu X, Zhang J, et al. 2021. Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science* **374:** abe6474. doi:10.1126/science.abe6474

Gene-Expression Profiling to Decipher Breast Cancer Inter- and Intratumor Heterogeneity

Alexander Swarbrick, Aranzazu Fernandez-Martinez and Charles M. Perou

Cold Spring Harb Perspect Med published online May 3, 2023

Subject Collection Breast Cancer: From Fundamental Biology to Therapeutic Strategies

For additional articles in this collection, see http://perspectivesinmedicine.cshlp.org/cgi/collection/