Ovarian cancer is a complex disease with multiple subtypes, each having distinct histopathologies and variable responses to treatment. This review highlights the technological milestones and the studies that have applied them to change our definitions of ovarian cancer. Over the past 50 years, technologies such as microarrays and next-generation sequencing have led to the discovery of molecular alterations that define each of the ovarian cancer subtypes and has enabled further subclassification of the most common subtype, high-grade serous ovarian cancer (HGSOC). Improvements in mutational profiling have provided valuable insight, such as the ubiquity of TP53 mutations in HGSOC tumors. However, the information derived from these technological advances has also revealed the immense heterogeneity of this disease, from variation between patients to compositional differences within single masses. In looking forward, the emerging technologies for single-cell and spatially resolved transcriptomics will allow us to better understand the cellular composition and structure of tumors and how these contribute to the molecular subtypes. Attempts to incorporate the complexities ovarian cancer has resulted in increasing sophistication of model systems, and the increased precision in molecular profiling of ovarian cancers has already led to the introduction of inhibitors of poly (ADP-ribose) polymerases as a new class of treatments for ovarian cancer with DNA repair deficiencies. Future endeavors to define increasingly accurate classification strategies for ovarian cancer subtypes will allow for confident prediction of disease progression and provide important insight into potentially targetable molecular mechanisms specific to each subtype.

Summary Sentence
Advances in technologies to define the transcriptional profile of ovarian cancers has resulted in continuing evolution of subtype classification, improvements in preclinical animal models and a greater appreciation for the complexity of these tumors.

Key words: ovarian cancer, subtype, transcription, microarray, single-cell RNA sequencing, animal models.
Figure 1. Summary of ovarian cancer subtypes. A chart summarizing ovarian cancer histological (A) and molecular (B) subtypes, their key features, and their relations to each other. Histological subtypes were originally defined by Serov et al. [8]. The molecular subtypes of serous ovarian cancer have been defined by Tothill et al. [32], TCGA, 2011 [2], and Tan et al., 2013 [37].

ovarian cancer (HGSOC) that correlate with clinical outcomes. With recent advances in single-cell and spatially resolved transcriptomics, we will soon have an understanding of how the cellular composition and structure of tumors contribute to these molecular subtypes. Accurate classification strategies for ovarian cancer subtypes allow for confident prediction of disease progression and provide insight into potentially targetable molecular mechanisms specific to each subtype.

In this review, we will highlight these technological milestones and the studies that have applied them to change our definitions of ovarian cancer. We will also discuss emerging technologies and predict how they may add to our increasingly complex model of ovarian cancer biology. Finally, we discuss the development of targeted therapies based on characteristics of individuals' tumors. This review primarily focuses on advances in transcriptional profiling and how they have influenced our classification of the disease, but it is important to note that improvements in mutational profiling have also provided valuable insight into its genetics, such as the ubiquity of TP53 mutations in HGSOC tumors [2]. This information has been critical for establishing relevant preclinical cancer models, which has been reviewed extensively [3–7]. Additionally, work in this area has revealed that beyond TP53 mutations, recurrent oncogenic mutations are rare, but the disease is associated with high genomic instability that gives rise to incredibly complex mutation signatures [8–10]. It is important to consider how this additional complexity contributes to ovarian cancer heterogeneity, but given its complexity, we believe it is deserving of its own review, and as such, we do not discuss this work in great detail.

Early standardization of ovarian cancer subtypes

In 1952, a subcommittee of the World Health Organization (WHO) Expert Committee on Health Statistics raised concern about the poorly standardized nomenclature of cancer biology [11]. Appropriate language had not been established to describe the disease, and identical pathologies were frequently described differently by various medical professionals. They believed that three separate classifications were required, according to the tumor’s anatomical site, histological type, and degree of malignancy. In 1956, the WHO passed a resolution to explore the possibility of organizing centers across the world for the collection and classification of human tumor samples, and to establish histological definitions of various cancer types. Twenty-three centers were established across the world from 1958 to 1973, each specialized for cancers of specific anatomical sites.

The WHO International Reference Centre for the Histological Classification of Ovarian Tumors was established in 1963 in the Research Institute of Oncology of the Academy of Medical Sciences of the USSR in Leningrad. About 10 years later, in 1973, they released their report defining histological subtypes of ovarian cancer [11]. The report outlines specific histological features of each subtype (Figure 1A). Serous ovarian cancer—the most common histological subtype—for example, is described as tumors composed of epithelial cells resembling the epithelium of either the fallopian tube or ovarian surface, and while ciliated cells are found almost always in benign tumors, they are rare in malignant tumors. The classifications of serous, endometrioid, clear cell, and mucinous epithelial ovarian tumors are still used today and much work has focused on identifying the unique biology and clinical features associated with each subtype [12–14].

Emergence of molecular classifications of ovarian cancer

Along with the development of current histological subtyping schemes, classic chromogenic stains used in histology began to shed light on the molecular composition of tumors. In the initial report, the WHO also presented several of these stains to help define subtypes, including mucicarmine to identify mucin in serous tumors,
Oil Red O for high lipid content in theca cell tumors, and reticulin to stain reticular fibers of granulosa and theca cell tumors [11]. While immunohistochemistry was first developed in 1941 [15], it was not until the early 1970s that the approach began to be used to interrogate the molecular features of ovarian tumors. These earliest studies sought to identify cancer-specific antigens that could be used to discriminate between normal and malignant tissue. Several groups demonstrated that antisera of rabbits exposed to lysate from human tumors could be used to stain ovarian cancer with little reactivity to healthy ovarian tissue [16, 17]. Problematically, however, the specific antigens providing this specificity were not known.

The lack of defined ovarian cancer–associated antigens persisted for nearly a decade. In 1981, Bast et al. [18] derived a hybridoma cell line from antisera of mice immunized to the ovarian cancer cell line OVCA433, and purified the first monoclonal antibody, OC125, with an apparent specificity to epithelial ovarian tumors, reacting to the antigen CA 125 (MUC16). The authors demonstrated that the antibody reacted with all ovarian cancer cell lines tested, and tumor tissue from 12/20 of the patients assessed. It also had no binding to a variety of non-malignant tissues, including fetal and adult ovary, or any sections from 12 non-ovarian carcinomas. Within 2 years, OC125 was adapted to measure CA 125 levels in serum and could be used to monitor the progression of ovarian tumors following treatment [19]. They showed that 99% of healthy individuals had CA 125 levels lower than 35 U per milliliter, while 82% of ovarian cancer patients had elevated levels of the antigen. Fluctuations in CA 125 levels also correlated with the progression/regression of the disease. This sparked a surge of work throughout the 1980s to validate the specificity and sensitivity of using CA 125 to diagnose and monitor epithelial ovarian cancer. Today, despite CA 125 fluctuations also being associated with non-gynecological malignancies (e.g. pancreatic cancer and other non-malignant conditions (e.g. inflammatory conditions in the abdomen)), its measurement remains an FDA-approved standard for monitoring therapy response in ovarian cancer patients and to detect residual or recurrent disease following first-line therapy (reviewed in Bottoni and Scatena [20]).

This work represented the first wave of molecular profiling ovarian cancer for diagnosis and surveillance based on the expression levels of a gene. In the decades since, many studies have begun to piece together distinct molecular characteristics of these tumors. The markers currently used are largely products of focused, hypothesis-driven studies, with candidate markers having been carefully selected based on previous knowledge. The expression of the Wilms’ tumor gene WT1—a gene commonly used as a marker of serous tumors—was originally explored in ovarian tumors because of its involvement in Wilms’ tumors of the kidney and in the development of genitourinary organs. Because of this, several groups assessed it in ovarian cancer and discovered that it was a reliable marker of serous tumors [21–23]. While this pick-and-choose approach has worked for developing several standard tests, its main limitation is that it depends on previous knowledge about candidate markers. As a result, many genes would never be explored as candidate markers because of a lack of prior evidence. Prioritizing candidate markers is also difficult because of the challenge of weighing pre-existing evidence. This approach, however, was a product of the technology available at the time: it was simply impossible to screen markers in an unbiased manner, so candidates had to be chosen carefully. It was not until the emergence of DNA microarrays in the mid-1990s that this began to change [24, 25].

**Microarrays and early signs of heterogeneity within histological subtypes**

Microarrays enabled gene expression analysis at unprecedented throughput. The expression of thousands of genes could be assessed in a single experiment, dramatically changing how experiments could be performed. Candidate markers no longer needed to be chosen based on prior knowledge, but rather could arise from the data itself. While the technology was superseded relatively quickly by next-generation sequencing, the unbiased approach to science that it introduced has become ubiquitous across many fields of biology.

The first microarrays to assess gene expression profiles of ovarian tumors was from Wang et al. in 1999 [26]. The authors assessed the relative expression of 5766 different genes between several different ovarian tumors and normal ovarian tissue. The tumors analyzed included two serous, two endometrioid, one poorly differentiated tumor, one mucinous, and one clear cell tumor. They did not make comparisons between subtypes, but rather, looked for differences broadly between ovarian tumors and normal ovarian tissues. With this approach, they identified changes in several differences between these tissues, including reduced levels of JUN, FOS, and VIM (Vimentin), as well as elevated levels of several epithelial markers (e.g. cytokeratin 8), the protease inhibitor HE4, and mesothelin, which had been previously described as an ovarian cancer marker [27]. The second microarray study followed a year later by Ono et al. [28]. In this study, the authors compared five serous tumors and four mucinous tumors with normal ovarian tissue. Similar to the previous study, they showed that tumors, regardless of subtype, had elevated RNA content for cytokeratins and HE4. They also compared the two histological subtypes to identify subtype-specific expression patterns. Reduced Vimentin levels, which had previously been shown by Wang et al. [26], were associated with the serous tumors, but not the mucinous samples, which uniquely had elevated levels of several metabolic genes (e.g. CKB and ALDH1B1). While these two studies presented the first unbiased look into expression differences associated with ovarian tumors, it was still unclear how to gain useful biological insight beyond simply listing genes with the largest changes across samples.

Soon after these two studies, Welsh et al. [29] published the first microarray study that used analysis methods resembling the common approaches used today, using unsupervised clustering to group samples and define gene sets with similar expression patterns. This approach has the benefit of not relying on user-defined classifications to group samples, which introduces inevitable bias, but rather, the clustering algorithms group samples together based on similarities in their gene expression profiles. They performed microarrays on 38 experimental samples, including 27 serous tumors, multiple normal ovarian tissue samples, several ovarian cancer cell lines, activated B cells, and HUVECs (human umbilical vein endothelial cells). Across these samples, they found five gene expression signatures: proliferative, pan-cancer, pan-normal, stromal, and ribosomal groups. These classifications were also related to clinically relevant features. For example, tumors with higher expression of the ribosomal gene set, which cluster close to normal tissue samples, represented tumors with a more-differentiated histology. Tumors with high expression of the proliferative gene set, however, were less-differentiated and were associated with a high malignant potential [29]. This study was the first to provide evidence of transcriptional heterogeneity within a single histological subtype of ovarian cancer that corresponded to clinical features of the disease.
In parallel with work assessing transcriptional heterogeneity, Shih and Kurman spent years assessing variations in the pathogenesis and behavior of various histological subtypes, along with the genetic aberrations associated with each [30]. In 2004, they proposed a dualistic subtype classification scheme complimentary to histological subtypes that focused on modes of tumorigenesis and the unique genetic aberrations contributing to the process, rather than the histological patterns of the tumors [30]. The terms—Type I and Type II—were used to define different paths of tumorigenesis rather than the resultant tumor histology itself. Type I tumors included low-grade tumors (e.g. low-grade serous, mucinous, endometrioid, clear cell) that develop from borderline tumors and have frequent KRAS or BRAF mutations. Type II tumors were defined as high-grade tumors with frequent TP53 mutations and poorly defined precursors, although precursor lesions called p53 signatures or serous tubal intraepithelial lesions (STICs; resulting from TP53 mutation) have since been identified in the distal fallopian tube [31, 32]. The broad classification proposed by Shih and Kurman was invaluable for reshaping how subtypes could be defined, serving as a turning point for subtyping schemes based on molecular data rather than simply histology. Recent studies of integrated molecular profiling have in turn provided insight into the concept that HGSOC may originate from fallopian tubes or ovaries [33, 34].

Large-scale genomics studies and the chase for well-defined molecular subtypes

The early microarray studies began to demonstrate that further stratification within individual histological subtypes may be resolved from molecular profiles, but did not include sample sizes large enough to sufficiently power this type of analysis. The notion of molecular subtypes of ovarian cancer was formalized several years later by Tothill et al. [35] following their effort to generate expression profiles of 285 serous and endometrioid tumors. Clustering of these profiles revealed six molecular subtypes of serous and endometrioid tumors, vaguely dubbed “C1” through “C6” (Figure 1B). Similar to the previous studies, the authors found subtypes driven by epithelial (C4), proliferative (C5), or stroma-associated (C1) expression signatures, and signatures associated with tumors of low malignant potential (C3 and C6). They also identified two novel molecular subtypes, including clusters with elevated immune (C2) and mesenchymal-associated expression patterns (C5). These molecular subtypes were also associated with different cellular composition, such as reduced immune cell infiltration in the mesenchymal subtype, and the expression signatures were also predictive of overall survival, with the stromal subtype having the worst prognosis.

This was the first study to thoroughly define molecular subtypes and compare clinical information between each, which, if implemented in clinical diagnosis, could provide personalized insight into each patient’s prognosis. The potential power of this approach fueled the development of The Cancer Genome Atlas (TCGA) in 2006, whose goal was to demonstrate that high-throughput genomics technologies could be used to produce biologically meaningful, and clinically relevant conclusions [36]. The program started with a three-year pilot, focusing on glioblastoma, lung, and ovarian cancers. Although the TCGA project came to an end in 2017, it ultimately profiled tumors from a remarkable 11,000 patients across 33 different tumor types [37].

Results from their ovarian cancer cohort were first published in 2011 [2]. The data included mRNA and miRNA expression profiles, promoter methylation, DNA copy number alterations, and exome sequencing of 489 high-grade serous ovarian adenocarcinomas. From gene expression data, they defined four molecular subtypes of HGSOC, which they termed “immunoreactive”, “differentiated”, “proliferative”, and “mesenchymal”, based on the genes associated with each subtype and previously used nomenclature from Torrill et al. [35] (Figure 1B). The immunoreactive cluster was defined by elevated cytokine expression, including CXCL11, CXCL12, and the receptor CXCR3. The proliferative cluster interestingly had lower levels of previously established ovarian cancer markers, such as MUC1 and MUC16 (CA 125), but had elevated expression of proliferation genes, including MCM2 and PCNA, whereas the differentiated samples lacked proliferation-associated expression, but had high expression of these ovarian cancer markers. The mesenchymal subtype had elevated expression of myofibroblast (e.g. FAP) and microvascular pericyte markers (e.g. ANGPTL2 and ANGPTL1). Interestingly, none of the subtypes were significantly associated with differences in overall survival, but the mesenchymal subtype trended towards worse outcomes [2]. With more samples, more clinical patterns may have been identified.

Exome sequencing of the TCGA cohort of HGSOC tumors indicated the most common mutation occurs is TP53, and is present in approximately 96% of tumors [2]. BRCA1/2 were the next most-frequently mutated genes, being affected in 22% of tumors [2]. Beyond this, there are few recurrent oncogenic patterns among ovarian tumors. Aggregating genes from signaling pathways did, however, reveal that several pathways with frequent alterations, including RB, PI3K/RAS, NOTCH, and FOXM1 signaling pathways [2]. Interestingly, the molecular subtypes defined by gene expression data did not enrich for particular mutation signatures.

While the TCGA classification scheme has been shown to stratify samples from independent cohorts [2, 38], there is still no general consensus about an optimal classification strategy. Several groups have since used different combinations of datasets and clustering strategy to establish their own subtyping scheme [38–42]. While different approaches are used, all identify similar subtypes (i.e. proliferative, differentiated, immunoreactive, and mesenchymal patterns). One particularly interesting study from Tan et al. [39] integrated data from 16 different studies and generated a classification scheme from the 1538 tumors, which has been the largest data set used and the most powered analysis to date. Their analysis resulted in five distinct clusters: Epi-A (epithelial), Epi-B, Stem-A, Stem-B, and Mes (mesenchymal) [39] (Figure 1B). While the nomenclature is different, these groupings are largely consistent with the previous classification schemes (Figure 1). Additionally, perhaps due to the increased power of their analysis, they found a significant association between the defined subtypes and patient outcome. The Stem-A (similar to the proliferative subtype) and Mes (mesenchymal) subtypes were associated with a worse outcome. Mes tumors enriched for advanced stage, metastatic tumors, whereas the Epi-A subtype included low malignant potential tumors. Stem-B was associated with other histological subtypes that were included in the classification, which likely explains the absence of an equivalent group in the TCGA subtypes, as it only included serious tumors.

Tumor composition confounds bulk expression profiles

Early in the efforts to establish molecular subtypes of ovarian cancer from microarray and RNA sequencing (RNA-seq) data it became...
Differences in tumor composition confound transcriptional profiles. A schematic demonstrating problems arising from bulk analysis of tumors (e.g. microarray, RNA-seq, Western blot). With these bulk methods, it is possible that true changes in gene expression within a cell type are masked by changes in the relative frequency of that cell type. In the schematic, tumor cells (brown) express lower levels of Gene A in “Tumor 1” but they are half as prevalent in “Tumor 2.” Consequently, Gene A appears unchanged with bulk methods. Additionally, changes in the frequency of immune (green) and stromal (orange) cells result in significant changes in the measured expression of their marker genes (Gene B and Gene C, respectively). Single-cell analysis methods (e.g. scRNA-seq, CyTOF) allow for the detection of gene expression within the individual cell types and are not confounded by differences in cell composition. They also allow for the measurement of the relative frequency of each cell type within tumors.

Clear that the single expression profiles being generated were at least in part driven by differences in the cellular composition of the tumors that were being processed. Ono et al. and Welsh et al. both discussed this possibility in their early microarray studies [28, 29]. Tothill et al. took this further and experimentally assessed the composition of the tumors from their cohort, demonstrating that their subtypes had variable tumor content, represented by the percentage of cells expressing the marker MUC1 [35]. They also demonstrated that their C1 (mesenchymal/stromal) cluster had a surprisingly low tumor content, with high levels of stromal activation, defined by levels of desmoplasia. In fact, stromal activation directly correlated with the C1-associated expression signature. Perhaps the most variable feature of the tumors measured was the degree of T-cell infiltration, with the C5 subtype (proliferative; Stem-A) having the fewest T cells within the tumor, and unsurprisingly, the immunoreactive C2 subtype having the highest amount of T-cell infiltration [35]. Similar differences in tumor composition between molecular subtypes have also been described by others [43, 44].

Ultimately, compositional differences between samples will always confound the bulk assays, like classic microarray or RNA-seq experiments. These techniques involve the dissociation and lysis of relatively large amounts of tissue in order to provide enough biological material for cDNA synthesis and input into the array or sequencer. As a result, the single molecular profile generated from an experiment is the product of all cell types within the tissue sample and their relative proportions. This makes the interpretation of these molecular profiles quite difficult: observed differences between subtypes does not necessarily mean the cancer cells themselves are any different; any intratumoral heterogeneity of the cancer cells is averaged out by the pooling of RNA during lysis; and while the expression profile may suggest the presence of various cell types, it is never directly tested (Figure 2).

Methods have been developed to help assess these compositional differences, with most using deconvolution of bulk expression profiles to predict the relative fractions of various cell types. CIBERSORT, for example, can use expression profiles from pure cell types (e.g. FACS-sorted immune cell types) to predict the relative proportion of each in the bulk profile [45]. Similarly, groups have used techniques such as laser dissection of tissue sections to enrich for regions of cancer cells prior to measuring gene expression. Tan et al. analyzed 78 laser-captured samples, as well as microarrays of 142 ovarian cancer cell lines to demonstrate that the cancer cells themselves exhibit intrinsic heterogeneity from sample to sample and can be classified in each of the five subtypes they had defined [39]. If the whole tissue were processed, however, it is unclear if it would yield the same subtype classification. Also, by assessing the cancer cells in isolation, relevant information about tumor composition is lost (e.g. immune cell infiltration).

This raises an interesting question: is it important—practically—that the expression profiles are confounded by cell composition? While it is difficult to argue that it does not hinder biological interpretation of the expression profiles, perhaps an argument could be made that these profiles could be sufficient for predicting clinical outcomes. Expression heterogeneity within the various cell types and differences in their proportions will both contribute to the bulk expression profile. These two sources of variation are likely both clinically relevant, so does it matter that we cannot tease apart the contribution of each to the bulk profile? With enough reference samples, perhaps the bulk profiles could capture all of the clinically relevant variation, even if it’s not clear what specific cell types contribute
to the profiles. This is supported by the studies that have shown that these profiles can, in fact, predict clinical outcomes [39].

While this argument is reasonably convincing, the underlying biology missed by this approach may be too important to ignore. With bulk assays, many sources of variation are compounded into a single measurement, with larger sources of variation (e.g., cell type composition) potentially masking clinically important, but less dramatic sources. Examples may include cancer stem cells which are a putative population of rare cells with high tumorigenic potential and an intrinsic resistance to chemotherapy [46]. Their existence in ovarian tumors is still unclear [47, 48], but if present, bulk approaches may not provide the necessary sensitivity to reliably detect them among the rest of the population. With bulk measurements, it is also possible for the detection of true gene expression changes within a cell type to be negated, or even reversed, by compositional differences—for example, when increased expression level of a gene within a specific cell type is coupled with a reduced representation of that cell type within a tissue (Figure 2). While bulk profiles from tumors can provide value, their interpretation is often challenging and should be approached with caution.

Single-cell genomics as a tool for untangling tumor complexity

In 2009, Tang et al. demonstrated the first protocol to perform RNA-seq on a single cell by manual isolation and processing of oocytes and blastomeres [49]. In the 10 years since, the field of single-cell genomics has exploded. The adoption of microfluidics has enabled new, efficient methods for producing cDNA libraries from large numbers of single cells. The most common approach is to use a microfluidics system to encapsulate single cells in aqueous droplets suspended in oil. Each droplet contains the reagents necessary to lyse the cell and drive cDNA synthesis. A DNA barcode unique to each droplet is added to all cDNA molecules in this reaction, allowing cDNA to be pooled and sequenced together. The sequencing data is demultiplexed by assigning each sequencing read to a cell (determined by the unique barcode sequence) and its corresponding gene [50, 51]. This has dramatically increased the throughput of the technique, making it feasible to perform genome-wide expression profiling on tens of thousands of cells in a single experiment. The commercialization of this approach has made the technology readily available and user-friendly, and single-cell RNA sequencing (scRNA-seq) has become a standard tool in the biologist’s arsenal, allowing for rapid identification of the different cell types present in tissues, and even heterogeneity across cells of an individual cell type (e.g., cell cycle state, apoptosis, hypoxic response, etc.).

At the time of writing this review, there has only been a single scRNA-seq ovarian cancer study published using this high-throughput approach. Shih et al. [52] generated single-cell libraries from 2911 cells collected from 14 samples, including matched primary and metastatic masses from low-grade and high-grade ovarian tumors. With only several hundred cells per sample, the study wasn’t sufficiently powered to make strong claims about individual samples or to detect rare cell types. Using the aggregated data, however, the authors were able to show large amounts of variation in the composition of tumors, with metastatic samples having a surprisingly low percentage of cancer cells (∼10%) compared to matched primary tumors (∼68%). Using gene sets associated with the four TCGA molecular subtypes, the authors also found that aggregating all cells from individual samples often did not result in strong classifications, but the individual cell types comprising the data could be readily classified. This supports that molecular subtypes defined from bulk data are likely driven by the composition of the tumor.

There has also been a single study that performed much lower-throughput scRNA-seq on ovarian cancer cells from a fresh HGSOC tumor [53]. Winterhoff et al. [53] collected 92 single cells from a HGSOC tumor sample that was depleted of immune cells. These cells fell into two large categories, seemingly corresponding to the epithelial (carcinoma; EPCAM+) and stromal components of the tumor (CD44+). There was also heterogeneity within each of these broad clusters, but this was not explored deeply. When the bulk profile from this tumor sample is classified with the TCGA’s four-subtype scheme, it is most consistent with a “mesenchymal” subtype. Interestingly, however, when the carcinoma and stromal components were classified independently, the carcinoma cells were consistent with a “proliferative” classification, while the stromal component were “mesenchymal”, further highlighting that tumor composition may be the feature driving subtype classifications, rather than the cell state of the cancer cells themselves [53].

Unfortunately, with the limited number of cells captured, and the depletion of immune cells during tissue processing, many complexities of the tumor remain unassessed. The authors searched for populations associated with chemoresistance and cancer stem cell expression signatures, but were unable to find consistent patterns, likely because of the rarity of cells expressing markers of these cells. A larger number of cells may be required to confidently identify these populations.

scRNA-seq is not the only high-throughput approach to assess heterogeneity at single-cell resolution. Mass cytometry (CyTOF) enables single-cell quantification of dozens of proteins in parallel using heavy metal-tagged antibodies, which are discernable due to their unique mass signatures in time-of-flight mass spectrometry [54]. Because it does not involve sequencing, it is trivial to measure a larger number of cells than with scRNA-seq. A limitation, however, is the low number of genes quantified (usually ∼30–40 proteins) compared to scRNA-seq experiments (genome-wide; ∼20,000). These techniques could ultimately become complimentary, with scRNA-seq being used to identify the genes with the most relevant variation on a small number of samples, and CyTOF being used to screen the given panel of informative genes across many samples.

Gonzalez et al. [55] took advantage of mass cytometry to assess heterogeneity of over 800 000 cells across 17 HGSOC samples. They used single antibodies to manually gate out immune cells (CD45+), endothelial cells (CD31+), and stromal cells (FAP+) from their data, but included a panel of 38 cancer-associated antibodies, including markers of proliferation, cell survival, the stem cell state, the epithelial-mesenchymal transition, and several more cellular functions. Lacking genome-wide expression data, they were unable to perform pre-existing subtype classification schemes, but were able to capture a large amount of heterogeneity among cancer cells, identifying 56 distinct clusters of expression profiles, with 10 being dominant clusters that occur at relatively high proportions across the 17 tumors. It should be noted that many of these clusters are driven by subtle variation and it is unclear how clinically relevant this amount of resolution truly is. The authors did, however, find that the epithelial-mesenchymal transition was a large source of variation across the cancer cell clusters, with clusters being strongly epithelial (CDH1+/VIM−), strongly mesenchymal (CDH1−/VIM+), or exhibiting a partial EMT state (CDH1+/VIM+). The relative proportion of a subcluster of the mesenchymal branch co-expressing VIM, MYC, and HE4 was found to be associated with carboplatin
resistance and tumor relapse, which is consistent with previous findings demonstrating that the “Mes” subtype is associated with worse overall survival [39]. Interestingly, this population was present in all tumors assessed, but at varying frequencies (1%–15%), where patients having tumors with the highest frequency fared worse. The presence of these cells in all patients highlights an interesting question of why relatively small differences in their frequency can impact patient outcome so dramatically. Additionally, it reinforces the importance of having strategies to detect these low-frequency, but clinically relevant differences, and so adopting single-cell strategies may be important to accomplish this.

High-throughput scRNA-seq and CyTOF will ultimately improve and reshape our understanding of ovarian cancer biology and its molecular subtypes, allowing scientists to measure tumor heterogeneity, unconfounded by the cellular composition of tumors. Similar to the early days of microarray analysis, it will likely not be possible to establish consistent subtypes of ovarian cancer until dozens or hundreds of scRNA-seq samples have been collected, but in the meantime, the data will provide valuable insight into tumor biology. Other cancers have already benefited from this approach. For glioblastoma, scRNA-seq has been used to show that individual tumors contain cells corresponding to multiple molecular subtypes previously defined by the glioblastoma TCGA cohort [36, 56]. scRNA-seq from several melanoma samples has demonstrated intratumoral heterogeneity associated with the presence of a dormant, drug-resistant phenotype, but the frequency of this population varies between patients [57]. In head and neck squamous cell carcinoma, although cancer cell expression is quite variable among patient samples, seven gene expression programs exhibiting intratumoral heterogeneity across all tumors were identified, including a partial EMT program that is predictive of metastasis [58]. In a high-throughput scRNA-seq study of over 50,000 cells from 15 lung cancer samples, the authors were able to catalog 52 different noncancer cell types present in the tumors and defined unique expression markers for each that could be assessed in bulk RNA-seq from the lung cancer TCGA data set and correlated with patient survival [59]. The adoption of these techniques for profiling ovarian cancer will address many outstanding questions about the disease. Are compositional differences of ovarian tumors, such as reduced immune cell infiltration, associated with expression changes in cancer cells? How much intratumoral heterogeneity do ovarian cancer cells display? Are there any reliably detectable populations that could correspond to rare cancer stem cells? Do histological patterns such as “differentiated” or “undifferentiated” overlook importance nuances that can be observed from the molecular profiles of single cells? How much do both composition and expression differences contribute to the molecular subtypes derived from bulk data? Will these molecular subtypes even be reproducible with single-cell data?

It will surely be exciting as this technology continues to scale towards hundreds of thousands, if not millions of cells per sample. Recent advances in multiplexing have also made it feasible to pool and process multiple samples together in single runs, improving the throughput and reducing the cost per sample [60–62]. Groups have also begun modifying the scRNA-seq methodology to allow for the read-out of multiple biological features from individual cells. Cell surface proteins on single cells can be quantified in parallel with RNA expression by exposing the cells to barcode-conjugated antibodies prior to processing, which produces flow-cytometry-like profiles in conjunction with gene expression data [63]. The combination of gene expression quantification along with T cell and B cell receptor sequencing has recently been commercialized, allowing for the identification of immune cell clonotypes within tumors, which can be used to assess clonal expansion of these cells, providing information about the immunogenicity of the tumor [64]. The combination of scRNA-seq and ATAC-seq from the same cell has also recently been reported [65], which may help determine transcription factor drivers of various cell populations with tumors. Others have also began combining pooled CRISPR library screens with scRNA-seq, enabling matching sgRNA sequences to genome-wide expression profiles from populations targeted with a library of CRISPR sequences [66, 67]. While this has yet to be applied to in vivo tumor models, it would be a valuable tool for assessing perturbations that enhance/restrict tumorigenesis or drug resistance in ovarian cancer models. Finally, whole-genome DNA sequencing has providing valuable information about ovarian tumor clonality and mechanisms of metastatic spread [68], but single-cell DNA sequencing is increasing in popularity and its increased resolution would enable study of rare clones and how they evolve throughout the progression and treatment of the disease.

**Structural features of ovarian tumors and spatially resolved analysis**

Pathologists often rely on visually detecting patterns in tissues to diagnose disease, such as the papillary structures of serous ovarian tumors, or the cystic glands of mucinous tumors. Assessing tumor structure has provided insight into the disease from the earliest days of oncology. Yet, as increasingly powerful quantitative methods to analyze these tissues have been developed, the methods have become more accurate through the structural and functional changes associated with tumor structure and organization. Because a cell’s gene expression program is a function of its surroundings, it is likely that spatial features are reflected in gene expression profiles of single cells, but these relationships are currently unknown, and thus spatial information cannot yet be inferred from expression data alone.

Despite being classified as a single histological subtype, the structure of HGSOC tumors can be remarkably diverse, even within the same patient [69–71]. Molecular subtyping of multiple collections from the same HGSOC tumors has demonstrated that these subtypes are most often not homogeneous throughout [42]. In one case, six collections from a single tumor yielded four of five possible molecular classifications [42]. Metastases often vary quite differently from the primary tumor, and a metastasis itself can exhibit subtype heterogeneity [42]. Interestingly, the C1 (mesenchymal) subtype, which is characterized by a more invasive phenotype, is often the primary tumor, and a metastasis itself can exhibit subtype heterogeneity [42]. Metastases often vary quite differently from the primary tumor, and a metastasis itself can exhibit subtype heterogeneity [42]. Interestingly, the C1 (mesenchymal) subtype, which is characterized by a more invasive phenotype, is often seen.

**Defining Ovarian Cancer Subtypes, 2019, Vol. 101, No. 3**

651

---

**Downloaded from** https://academic.oup.com/biolreprod/article-abstract/101/3/645/5513989 by University of Ottawa user on 29 December 2019
activation, preventing an immune response and worsening patient outcome [75]. A contradictory study found PD-L1 expression to be associated with better outcome in a cohort of ovarian cancer patients, and found macrophages to be the major source of its expression in these patients [76]. In melanoma, however, cancer cells adjacent to infiltrated lymphocytes exhibit elevated expression of PD-L1, inhibiting the immune response regionally [77]. Therefore, it is likely that the relevance of cell types present in tumors and their transcriptional state is likely dependent on their structural organization and interactions within the tumor.

Recent studies have also begun to search for conserved structural patterns across the TME of ovarian tumors and have attempted to relate these patterns to molecular subtypes. Zhang et al. identified three structural patterns associated with tumor infiltrating lymphocytes (TILs): N-TIL (sparse lymphocyte infiltration), S-TIL (lymphocyte infiltration restricted to stromal regions), and ES-TIL (lymphocytes infiltrate both stromal and epithelial nests) [71]. These structural patterns enriched for particular molecular subtypes, but the association was not perfect. Generally, N-TIL was associated with C4 and C5 (differentiated and proliferative, respectively), S-TIL was often found in C1 tumors (mesenchymal), and C2 tumors (immunoreactive) often contained ES-TIL structures. By assessing multiple samples from each patient in their cohort, the authors also found over 50% of the patients harbored tumors with multiple TIL subtypes [70]. TIL patterns also change throughout the course of therapy, with increased infiltration being associated with an immunogenic response following platinum-based treatment [69, 71]. Given the importance of lymphocyte infiltration in the progression of ovarian cancer, these structural patterns are important to consider, and it may be informative for predicting the efficacy of new immunotherapies, including immune checkpoint inhibition [78, 79].

It is evident that while both molecular subtyping and structural subtyping (e.g. TIL patterns) can be valuable for understanding tumor biology and predicting patient outcome, the two methods are mutually exclusive. Because of the documented intratumoral heterogeneity for both structural and transcriptional patterns, it cannot be assumed that a piece of tissue used for gene expression analysis has the same properties as the adjacent piece of tissue that was sectioned for histology. Methods that allow for simultaneous identification of transcriptional and structural patterns will enable integration of these approaches for, hopefully, more predictive subtyping. While immunohistochemistry has allowed for spatially resolved, single-cell resolution detection of protein expression, the number of genes that can be detected in parallel is limited by the number of unique stains that can be imaged. So, while cellular organization can be assessed, deep information about the transcriptional state of each cell cannot. Recent techniques have begun to emerge that allow for spatially resolved, high-throughput gene expression analysis.

The first approach that began to show promise towards this goal was multiplexed error-robust fluorescence in situ hybridization (MERFISH) [80, 81]. The method is an advancement of existing single-molecule fluorescence in situ hybridization (smFISH) approaches, which use fluorophore-labeled nucleic acid probes that are complementary to RNA species of interest. After cells have been fixed and permeabilized, these probes can be added to quantify the number of transcripts for a specific RNA species in each cell at single-molecule resolution [82]. MERFISH expands the technology by allowing for the quantification of hundreds-to-thousands of different RNAs. It accomplishes this by first coding all transcripts of interest as N-bit binary words, where N is the number of labeling iterations to be performed (e.g. Gene 1: “01”; Gene 2: “10”; Gene 3: “11”), and after a sequential labeling strategy of pooled FISH probes, each transcript in the cell will have produced a unique binary staining pattern (e.g. a transcript corresponding to Gene 3 in a cell would have yielded a positive signal after both rounds of labeling, whereas Gene 1 would be negative the first round, and positive the second; Figure 3A). This approach is promising because it provides spatially conserved information at truly single-cell resolution. It is, however, not genome-wide expression analysis, but the ability to scale the technique up to ~1000 genes may be sufficient for capturing relevant expression patterns. This method has yet to be used to assess tumor samples and has not been widely adopted, likely due to the requirement of relatively complex imaging equipment, and image analysis expertise.

Another approach holds the promise of being easier to adopt for most cancer biology labs, as it is based on techniques already commonly used: tissue sectioning and next-generation sequencing. “Spatial Transcriptomics” (ST) was developed in 2016 and allows for spatially resolved RNA-seq from fresh frozen tissue sections [83]. The method involves taking a section of tissue and permeabilizing it onto a glass slide that has been spotted with an array of mRNA-capturing probes, where probes from each spot contain a unique DNA sequence that serves as a spatial ID, providing spatial information for each sequencing read (Figure 3B). The spots of capture probes are 100 μm in diameter and 200 μm from center to center, capturing the transcriptome of ~10–20 cells. While not truly single-cell resolution, it provides global expression analysis throughout the tissue. Just as bulk tumor RNA-seq can be deconvolved to estimate the proportion of cell types within it, data from each spot of the ST array can also be deconvolved to predict the cell types of the 10–20 cells per spot [84]. This approach has been used to assess structural features of pancreatic ductal adenocarcinoma, where the authors used scRNA-seq to produce pure cell type expression profiles that served as references for deconvolution of each spot on the ST array [84]. As a proof-of-concept, the authors clustered the data associated with each spot and revealed structural patterns throughout the tissue, such as nest of cancer cells infiltrated by T cells adjacent to healthy ductal cell-rich tissue lacking lymphocyte infiltration [84]. This approach has also been used with serial sections of HER2+ breast cancer samples, generating an immune score across each section, and allowing for three-dimensional rendering of immune cell-rich niches within the tumor [85].

Similar to the issues of scRNA-seq studies, current studies have not processed enough samples to begin generating new subtype classification schemes, but these early experiments provide promising evidence that we will be able to reconcile structural patterns with molecular heterogeneity. Relationships between cancer and immune cell proximity within ovarian TME and their transcriptional state will be revealed, enabling a better understanding of cancer cells’ adaptive resistance in response to immunologic threat. It will also be valuable for assessing the intratumoral variation of these patterns. Eventually, with enough samples, these integrated patterns can be associated with various clinical features of interest, including metastatic progression and therapeutic response.

### Hopeless complexity: the curse of dimensionality

As we continue to pursue our ambitions of “the quantified tumor”, measuring the gene expression profile of all cells within a tumor and their spatial relationship to each other—not to mention the genome-wide mutation and epigenetic profiles of each cell—we are faced with immense complexity. Consider scRNA-seq data alone:
quantifying 20 000 genes across 10 000 cells from a single tumor already generates 200 million measurements. While adding more and more variables to our measurements gets us closer to capturing an unconfounded representation of cancer, the data becomes increasingly difficult to interpret. This is not simply due to the sheer volume of data, but also because of the well-documented “curse of dimensionality”: as more variables are measured, clusters of samples become increasingly less defined and each sample begins to look more and more unique. A more-mathematical representation of this highlights the problem nicely: as the number of dimensions (i.e. variables) for a given set of data increases towards infinity, the difference between the distance from a point to its nearest and furthest neighbors approaches zero. This is problematic because defining cancer subtypes depends on identifying clusters of samples that are similar to each other, but different from others. So, is our unfortunate fate that we will discover that all tumors are unique and subtypes don’t exist?

Fortunately, this is doubtful. While we are measuring an increasing number of variables, there is a high degree of informational redundancy. For example, with gene expression data, gene coregulation produces massive amounts of redundancy (in terms of information, not biochemical function), where a single transcription factor may drive the expression of hundreds/thousands of genes, and the expression of that entire coregulated module could be represented as a single variable (e.g. T cell activation). In fact, principal component analysis of scRNA-seq tumor data can often capture the majority of transcriptional variation within ~30 dimensions rather than the 20 000 dimensions (i.e. genes) quantified. This suggests that the intrinsic dimensionality—not necessarily what we measure, but rather the number of independent sources of variation—is relatively small. Unfortunately, we don’t currently know all of these latent sources of variation in tumor biology or how they contribute to tumor progression, and we don’t know what minimum set of information can be measured to reasonably capture them (e.g. marker gene panels). Their evidence is, however, apparent from existing studies.

In the TCGA study, the subtypes defined were surprisingly not associated with differences in overall or progression-free survival [2]. The authors did, however, construct a 193-gene signature that was significantly associated with overall survival across multiple datasets [35, 86]. Several years later in a follow-up study, they used a larger number of expression profiles to refine this list and developed a gene signature named “Classification of Ovarian Cancer” (CLOVAR) comprising 100 genes that were the most correlated ($n = 47$) or inversely correlated ($n = 53$) with survival [43]. This signature was not associated with any of the four molecular subtypes they had previously defined. In fact, tumors from each subtype could be stratified into high- or low-risk groups based on their expression of the survival gene set [43]. This suggests that the prognostic gene signatures they established capture some latent variable that is largely independent from the variation captured by the molecular subtyping gene sets. As each may contribute to different features that are clinically relevant, we may have to adopt a multi-leveled subtyping scheme to optimally predict cancer progression.

Another trap we often fall into is shoehorning data into clean, discrete categories. Can tumors truly be stratified into discrete subtypes, or is the true nature more consistent of a continuum of states? Currently this is not clear. With the TCGA classification scheme, although one of the molecular subtype gene sets is dominantly expressed, the extent of gene set activation varies within each subtype [43]. When gene set activation levels are binarized, 82% of the tumors are positive for more than one subtype, with each gene
set having variable associations with each other. For example, the “proliferative” and “immunoreactive” subtypes are strongly anti-correlated \((r = -0.81)\), whereas the “immunoreactive” and “differentiated” subtypes having no relationship at all \((r = 0.00)\) [43]. This suggests that the intrinsic states of these tumors cannot simply be represented as four mutually exclusive states. Rather they can blur into each other and co-occur. This could be due to intratumoral structural heterogeneity, or it is possible that the gene sets contributing to each subtype are independently regulated and can co-occur in the same cells. In the case of the latter, the molecular subtypes themselves therefore may represent different latent variables. Regardless, information is lost by forcing each sample into a single classification (Figure 4). In the future, as targeted therapies against these molecular subtypes are developed, this may prevent patients who may respond favorably from getting treatment because aspects of their tumor were ignored.

Facing the complexity of ovarian cancer can be daunting. It is difficult to fight the disease with confidence when we know our understanding of it is only a sliver of its true nature. But there is comfort in the quote from the late British statistician George E. P. Box, “all models are wrong, but some are useful”. Even if our current models of ovarian cancer are a gross oversimplification, they have provided clinically valuable information. As our models become increasingly robust, hopefully so will their value (Figure 4).

**The search for targeted therapies of ovarian cancer**

Perhaps the most promising aspect of establishing comprehensive molecular subtypes of ovarian cancer is developing effective targeted therapies based on the specific molecular mechanisms driving each subtype. Breast cancer has benefited from targeted therapies since the 1970s, with the development of tamoxifen treatment for estrogen receptor-positive breast cancer, but despite decades of active research, treatment options for ovarian cancer are often limited to surgical resection and treatment with platinum-based chemotherapy [87].

Currently generating the most excitement in the field is the use of inhibitors of poly (ADP-ribose) polymerase (PARP) in patients with BRCA1/2 mutations. The PARP1/2 enzymes are involved in the DNA damage response and their inhibition in tumors lacking BRCA1/2, which are also involved in the DNA damage response, results in synthetic lethality [88, 89]. As a result, PARP inhibition has quickly become a targeted therapy for tumors harboring mutated BRCA1/2. For HGSOC tumors, PARP inhibition is consistently associated with a prolonged progression-free survival in patients with BRCA1/2 mutations, but patients often relapse and do not survive longer [90–92]. It is currently unclear why patients so consistently develop resistance to the treatment, but reversion mutations that restore the BRCA1/2 open reading frame has been shown to be associated with acquired resistance in these patients [93].

Interestingly, the benefit of PARP inhibition in patients lacking BRCA1/2 mutations is much more variable than those with mutations, with some patients experiencing benefits equal to those with BRCA1/2 mutation carriers [90]. Patients with mutations in other components of DNA repair pathways, such as RAD51C, RAD51D, and PALB2 (6–10%) [2] may also benefit from PARP inhibition for the same reasons. The promoters of BRCA1/2 are also often hypermethylated in HGSC, silencing their expression in these tumors [2]. A recent study demonstrated that homozygous, but not heterozygous methylation of the BRCA1 promoter predicted response of HGSC tumors to PARP inhibitors [94]. This suggests that mutations themselves are not necessary, but rather that PARP inhibition may be a viable therapy for any patient with genetic, epigenetic, or transcriptional changes that impair DNA repair by homologous recombination.

While BRCA1/2 mutations are present in approximately 22% of cases, PARP inhibition may be beneficial for an additional 25% of
HGSOC patients with impaired DNA repair [94]. However, identifying these patients requires comprehensive screening. While targeted screening of mutations in DNA repair genes could identify a subset of candidate patients, it would not work for patients with non-genetic causes of impaired DNA repair. Sensitivity associated with epigenetic silencing depends on homozygous methylation, but methylation assays on tumor tissue are difficult to confidently quantify. Transcriptional profiling may provide a consistent method for screening all patients. In fact, epigenetic silencing of BRCA1/2 produces a transcriptional program similar to BRCA1/2-mutated HGSOC [94, 95]. It therefore seems likely that it may also be effective at identifying other mutational or epigenetic changes that promote defects in DNA repair by homologous recombination.

While the potential of PARP inhibitors may present new options for a large number of patients, it is important to continue searching for new therapeutic avenues. Following preclinical studies demonstrating its role in ovarian tumor progression [96, 97], inhibition of vascular endothelial growth factor (VEGF) only resulted in modest improvements to progression-free survival, with no effect on overall survival when used in combination with standard treatment [98–100]. Several clinical trials have also been performed to assess the efficacy of inhibitors of the epidermal growth factor receptor (EGFR) and its family member HER2, but have had similarly disappointing conclusions (reviewed in Wilken et al. [101]), despite HER2 expression being associated with poor outcome in HGSOC tumors [102].

No clinical trials have yet to been performed based on targets identified from the transcriptional profiles of the molecular subtypes themselves. These profiles may, however, provide insight into subtype-specific targets. In their study to define new molecular subtypes, Tan et al. identified elevated expression microtubule/tubulin-related pathways in the Stem-A compared to other subtypes, and demonstrated their selective sensitivity to microtubule assembly inhibitors [39]. In a follow-up study, they used a similar approach and screened for kinases elevated in the Mes subtype [103]. They found that the receptor tyrosine kinase AXL was upregulated specifically in the subtype, and its inhibition in Mes ovarian cancer cells reduces their motility and growth in vitro, and abolishes their tumor formation in xenograft models [103]. This work highlights the promise of using molecular subtyping to identify new targeted therapies that could ultimately improve the prognosis of the disease.

**Hopeful strides: looking towards the future**

Through the remarkable advances in technology over the last 50 years, we have gone from studying macroscopic gross anatomy to assessing the expression of over 20 000 genes across thousands of single cells. Along the way we have constructed tumor classification schemes that can be used to predict outcomes of each patient’s disease. While our existing approaches are gaining oversimplifications of the complexity of ovarian cancer, they still hold empirical value.

But what do the next decades hold for us? As we continue to set into this current data-rich era, we must continue to look forward. Single-cell genomics will inevitably shed light on ovarian cancer biology that we have yet to appreciate, but as we dive further into “the quantified tumor”, we can’t forget that cancer is a physical entity. Like the physical objects around us, they are likely more than simply the sum of their parts, and so spatially resolved assays will help give context to the data we collect and will reveal patterns that could not be easily observed in the data alone. We are also victims of sampling bias, analyzing and processing only the small pieces of tissues we can physically (and analytically) manage. But just as spatial information cannot be forgotten about when analyzing quantitative data, global tumor structure and its location in the body should not be ignored when analyzing small sections of tissue. As spatial technology improves, it is likely that we will be able to process larger and larger pieces of tissue, but it does not seem likely with that we will be able to scale to the entire tissue with technologies as they look today. And even still, we would still lack information about how a tumor was precisely interacting with the rest of the patient’s tissues. Unfortunately, the solution to this problem is hard to imagine when considering current technologies. Perhaps advanced imaging technologies will one day allow us to pry into the molecular profiles of living patients’ tissues at single-cell resolution in situ. Or perhaps our mathematical models of biological systems will advance to the point where the entire complexity of the disease could be simulated from a limited set of input information that is easily attainable from a patient.

It is unclear what the next 50 years will hold, but we will continue to stride forward, facing ovarian cancer’s remarkable complexity along the way. Our models will improve and so will our ability to understand the disease. Subtype classification schemes will become more comprehensive, capable of predicting a wide range of clinical features about a patient’s cancer. Our models will reveal linchpins of the tumor’s biology that will allow us to apply therapeutic approaches that works best for that specific patient. And hopefully, 50 years from now, we won’t have to report 150 000 ovarian cancer-associated deaths each year [104].

**References**


mensional transcriptomics provides detailed information about immune cell distribution and identity in HER2 breast tumors. bioRxiv 2018; 358937.


Dr. Barbara Vanderhyden is a Professor of Cellular and Molecular Medicine at the University of Ottawa and a Senior Scientist in the Cancer Therapeutics Program at the Ottawa Hospital Research Institute. As the inaugural Corinne Boyer Chair in Ovarian Cancer Research, her research focuses on the factors involved in the initiation of ovarian cancer, and the generation of models that shed light on cancer susceptibility, tumor onset and progression. These models are also used in preclinical trials for the evaluation of novel therapeutic approaches, including targeted and immune therapies. In addition to developing new experimental models, Dr. Vanderhyden enjoys incorporating the latest technologies into her research, most recently the use of single cell RNA sequencing and MULTI-Seq to explore the phenotypic plasticity of epithelial cells. Many of these newer technologies are driven by the creative talents of trainees, such as the co-author of this review, David Cook.