

The Repertoire of Serous Ovarian Cancer Non-genetic Heterogeneity Revealed by Single-Cell Sequencing of Normal Fallopian Tube Epithelial Cells

Highlights

- The projection of FTE subtypes refines the molecular classification of SOC
- Comprehensive single-cell profiling of FTE cells identifies six molecular subtypes
- Substantial non-genetic heterogeneity of HGSOC identified in 1,700 tumors
- A mesenchymal-high HGSOC subtype is robustly correlated with poor prognosis

Authors

Zhiyuan Hu, Mara Artibani,
Abdulkhalil Alsaadi, ...,
Tatjana Sauka-Spengler,
Christopher Yau,
Ahmed Ashour Ahmed

Correspondence

christopher.yau@manchester.ac.uk
(C.Y.),
ahmed.ahmed@wrh.ox.ac.uk (A.A.A.)

In Brief

Using single-cell RNA sequencing, Hu et al. identify six subtypes of fallopian tube epithelium (FTE) cells in normal human fallopian tube tissues. The FTE cellular subtypes reveal intra-tumoral heterogeneity in serous ovarian cancer (SOC) and define SOC subtypes that correlate with patient prognosis.



The Repertoire of Serous Ovarian Cancer Non-genetic Heterogeneity Revealed by Single-Cell Sequencing of Normal Fallopian Tube Epithelial Cells

Zhiyuan Hu,^{1,2,3} Mara Artibani,^{1,2,4,15} Abdulkhalil Alsaadi,^{1,2,15} Nina Wietek,^{1,2,5} Matteo Morotti,^{1,2,5} Tingyan Shi,^{1,2} Zhe Zhong,^{1,2} Laura Santana Gonzalez,^{1,2} Salma El-Sahhar,^{1,2} Mohammad KaramiNejadRanjbar,^{1,2} Garry Mallett,^{1,2} Yun Feng,³ Kenta Masuda,^{1,2} Yiyian Zheng,^{1,2} Kay Chong,^{1,2} Stephen Damato,⁶ Sunanda Dhar,⁶ Leticia Campo,⁷ Riccardo Garruto Campanile,⁵ Hooman Soleymani majd,⁵ Vikram Rai,⁸ David Maldonado-Perez,^{9,10} Stephanie Jones,⁹ Vincenzo Cerundolo,¹¹ Tatjana Sauka-Spengler,⁴ Christopher Yau,^{3,12,13,14,*} and Ahmed Ashour Ahmed^{1,2,5,16,*}

¹Ovarian Cancer Cell Laboratory, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK

²Nuffield Department of Women's & Reproductive Health, University of Oxford, Oxford OX3 9DU, UK

³Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7BN, UK

⁴Gene Regulatory Networks in Development and Disease Laboratory, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 9DS, UK

⁵Department of Gynecological Oncology, Churchill Hospital, Oxford University Hospitals, Oxford OX3 7LE, UK

⁶Department of Histopathology, Oxford University Hospitals, Oxford OX3 9DU, UK

⁷Department of Oncology, University of Oxford, Oxford OX3 7DQ, UK

⁸Department of Gynaecology, Oxford University Hospitals NHS Trust, Oxford OX3 9DU, UK

⁹Oxford Radcliffe Biobank, Nuffield Department of Surgical Sciences, University of Oxford, Oxford OX3 9DU, UK

¹⁰NIHR Oxford Biomedical Research Centre, Second Floor, Unipart House Business Centre, Oxford OX4 2PG, UK

¹¹Human Immunology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK

¹²Centre for Computational Biology, Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham B15 2TT, UK

¹³Division of Informatics, Imaging and Data Sciences, Faculty of Biology Medicine and Health, The University of Manchester, Manchester M13 9PT, UK

¹⁴Alan Turing Institute, London NW1 2DB, UK

¹⁵These authors contributed equally

¹⁶Lead Contact

*Correspondence: christopher.yau@manchester.ac.uk (C.Y.), ahmed.ahmed@wrh.ox.ac.uk (A.A.A.)

<https://doi.org/10.1016/j.ccel.2020.01.003>

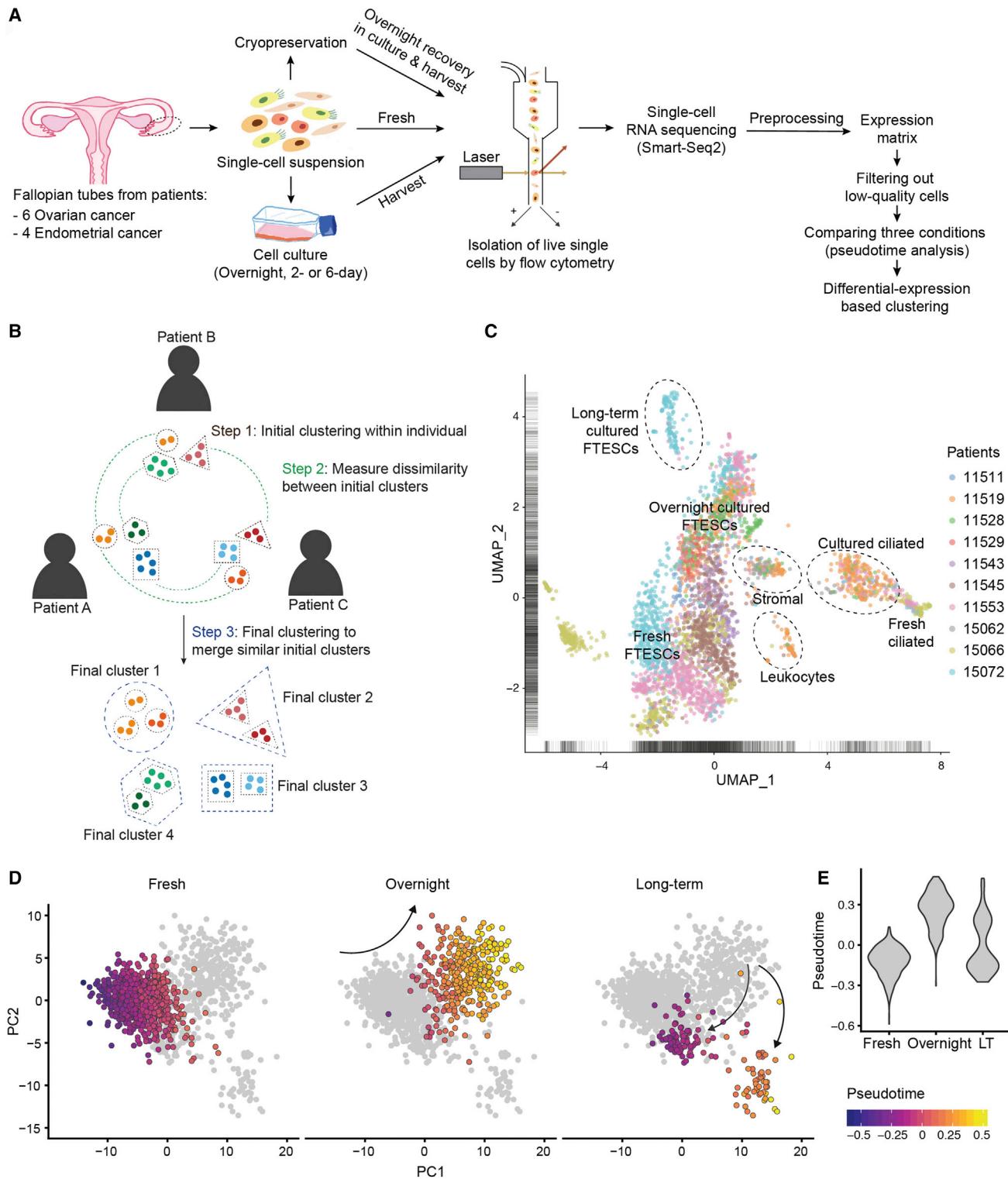
SUMMARY

The inter-differentiation between cell states promotes cancer cell survival under stress and fosters non-genetic heterogeneity (NGH). NGH is, therefore, a surrogate of tumor resilience but its quantification is confounded by genetic heterogeneity. Here we show that NGH in serous ovarian cancer (SOC) can be accurately measured when informed by the molecular signatures of the normal fallopian tube epithelium (FTE) cells, the cells of origin of SOC. Surveying the transcriptomes of ~6,000 FTE cells, predominantly from non-ovarian cancer patients, identified 6 FTE subtypes. We used subtype signatures to deconvolute SOC expression data and found substantial intra-tumor NGH. Importantly, NGH-based stratification of ~1,700 tumors robustly correlated with survival. Our findings lay the foundation for accurate prognostic and therapeutic stratification of SOC.

Significance

Serous ovarian cancers represent the most common malignancies of ovarian carcinomas. Molecular stratification of high-grade serous ovarian cancer (HGSOC) has been difficult, presumably because of profound genetic heterogeneity limiting reproducible prognostic classifications. We now overcome this limitation by investigating the differentiation trajectory of non-cancer FTE cells using single-cell RNA sequencing and discovering four FTE secretory subtypes. Using the subtype molecular markers of non-cancer cells, we define a gene signature that robustly identifies a poor prognosis EMT-high subtype of HGSOC. Conceptually, we show how examining non-cancer cells of origin could enable accurate prediction of cancer behavior. Moreover, our work has important prognostic implications for ovarian cancer patients and enables future efforts for therapeutic optimization for EMT-high tumors.



**Figure 1. A Cell Census of Human Fallopian Tubes**

(A) Diagram shows the single-cell RNA sequencing (scRNA-seq) and analysis workflow. We collected and processed normal fallopian tube tissues from ten cancer patients. All cells (directly sorted or maintained then sorted) were processed by using the Smart-Seq2 protocol. After the initial filtering, there were 3,877 good-quality cells left for downstream analysis. We compared the cells from three conditions to select the optimal condition for scRNA-seq. Cells from cultured or cryopreserved conditions, as well as cells carrying copy number variations and non-epithelial cells, were filtered out, which left 2,132 fresh FTE cells. Next, we used differential expression-based clustering to identify secretory subtypes.

(legend continued on next page)

INTRODUCTION

Intra-tumor heterogeneity is a key mechanism for survival and evolution of tumors (Brock et al., 2009; Marusyk et al., 2012). Genetic heterogeneity has been established as a mechanism of survival for most cancer types (Greaves and Maley, 2012; McGranahan and Swanton, 2017; Nowell, 1976). In addition, cancer cells of the same genetic background may have different phenotypic cell states that enable essential tumor characteristics, such as invasion, metastasis, and resistance to chemotherapy (Brock et al., 2009; Pisco et al., 2013). The ability of cancer cells to change from one cell state to another (plasticity) is a key feature for cancer survival (Meacham and Morrison, 2013). Although genetic heterogeneity is acquired, phenotypic heterogeneity is often inherited from the parent cell-of-origin of a tumor, due to epigenetic regulation or other mechanisms (Gupta et al., 2011; Visvader, 2011). Understanding cancer cell plasticity is dependent on the accurate identification and characterization of individual cell states. Recognizing that cancer cells have plasticity should be reflected on tumor stratification strategies that take into account the co-existence of multiple cancer cell states within any one tumor. However, the direct molecular characterization and identification of such states in a tumor is significantly confounded by genetic heterogeneity. An alternative approach for elucidating the phenotypic repertoire of a tumor type is to study the cell states of its cell-of-origin and use those characteristics to decompose an individual tumor into its constituents (Baron et al., 2016; Gupta et al., 2011; Newman et al., 2015).

High-grade serous ovarian carcinoma (HGSOC) is the most aggressive gynecological malignancy (Koshiyama et al., 2017), which is characterized by ubiquitous TP53 mutations and frequent chromosomal alterations (Ahmed et al., 2010; Bell et al., 2011; Etemadmoghadam et al., 2009; Macintyre et al., 2018). One of the major challenges of HGSOC is the late presentation with almost 80% of patients diagnosed at stage III or IV disease and a 5-year survival rate of about 30% (Torre et al., 2018). The absence of robust methods to stratify HGSOC has been a challenge against therapeutic innovation. Previous studies demonstrated the transcriptomic heterogeneity of HGSOC (Bell et al., 2011; Tothill et al., 2008). However, analyses using bulk transcriptomes are confounded by a variety of factors, such as copy number variation (Macintyre et al., 2018) and infiltration by non-cancer cells. Importantly, such analyses do not take into consideration the possibility of multiple cell states co-existing in one tumor. Such confounding factors lead to unstable classifications with variable prognostic power across independent datasets (Chen et al., 2018). It is plausible to hypothesize that phenotypic states of the cell-of-origin of ovarian cancer may be echoed in daughter cancer cells. Linking basal

and luminal cells from the mammary epithelium to corresponding breast cancer subtypes is a clear example of achieving a stable molecular classification based on understanding phenotypic diversity of the cell-of-origin. This approach has also been successful in other cancers (Fessler and Medema, 2016; Gilbertson, 2011; Ince et al., 2007; Wang et al., 2013).

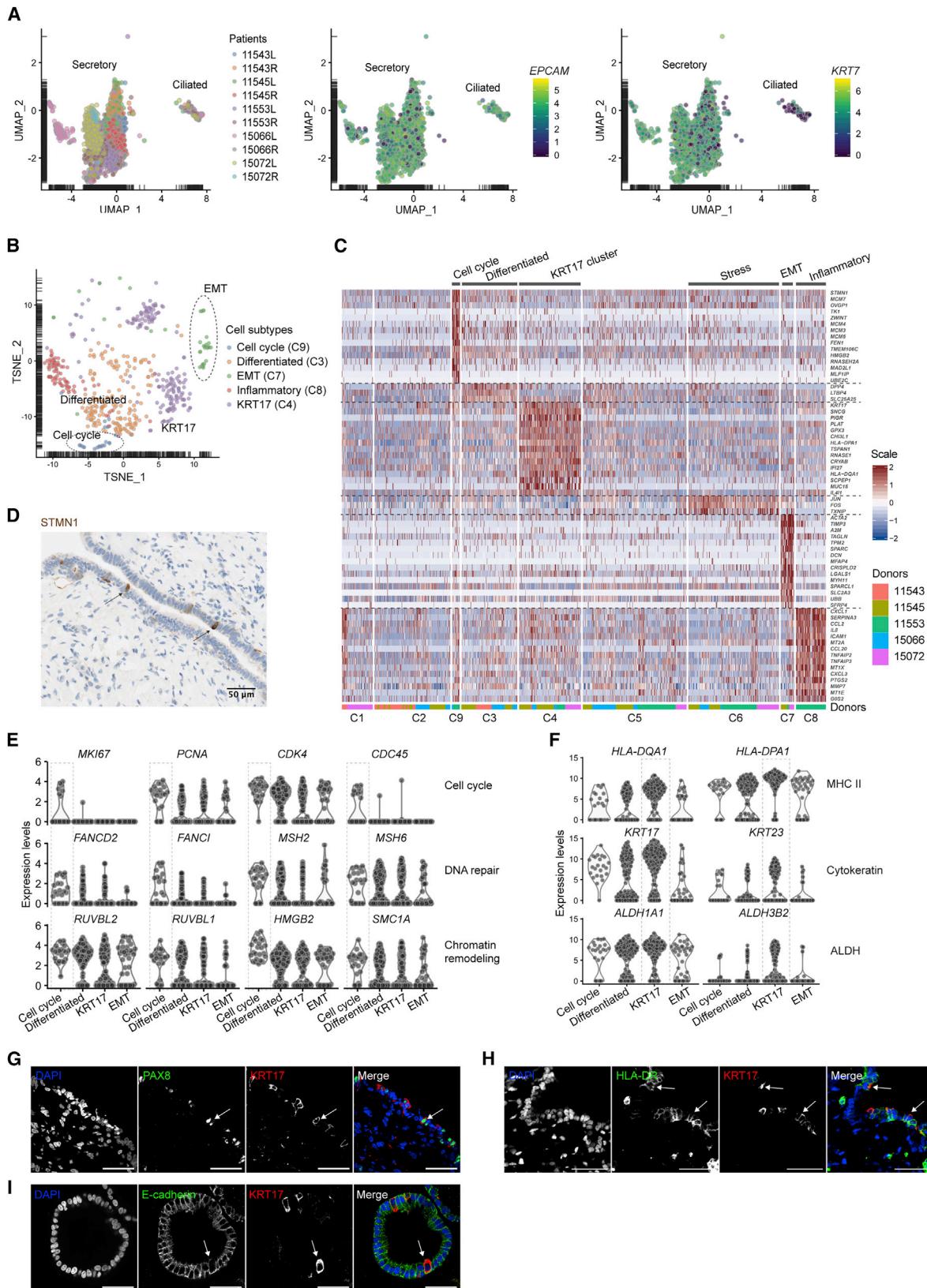
Recent studies strongly support that SOC can originate from the fallopian tube epithelium (FTE) with evidence from mouse models and genetic evolutionary studies (Ducie et al., 2017; Kim et al., 2012; Labidi-Galy et al., 2017; Perets et al., 2013). Previous work reported that the FTE is composed of PAX8-positive secretory, TUBB4-positive ciliated, and CD44-positive basal cells (Clyman, 1966). However, whether there are additional cellular subtypes and whether these subtypes are connected to subtypes of HGSOC have remained elusive. The advent of single-cell RNA sequencing (scRNA-seq) has enabled the identification of cellular subtypes (Grün et al., 2015). There are two major types of scRNA-seq techniques (Svensson et al., 2018): (1) the droplet-based methods that have the advantage of high throughput (from 10 to 100,000 cells), such as InDrop (Klein et al., 2015), Drop-Seq (Macosko et al., 2015), and the commercially available 10x Genomics Chromium system, and (2) plate-based methods that require fewer cells and have higher gene coverage, such as Smart-Seq2 (Picelli et al., 2014; Ziegenhain et al., 2017). Recent studies suggest that higher sequencing depth can compensate for a smaller sample size (Seirup et al., 2019; Svensson et al., 2019), and two studies have demonstrated that Smart-Seq2 can robustly identify cell states with relatively small sample sizes. Grün et al. (2015) sequenced only 238 cells from mouse intestinal organoids with the coverage of 3,000 genes per cell, which enabled the identification of rare intestinal cell types; Patel et al. (2014) analyzed 430 cells from 5 human glioblastoma samples with the average coverage of 6,000 genes per cell, which revealed the structure of transcriptional programs and cell phenotypes. Here, we characterized the cell subtypes of human FTE using this deep scRNA-seq technique, and investigated the association between the composition of bulk SOCs and prognosis.

RESULTS

Culturing Substantially Alters Single-Cell Transcriptomes

We first analyzed 3,877 single cells from the fallopian tubes of 5 ovarian cancer patients and 5 endometrial cancer patients using the Smart-Seq2 technique (Picelli et al., 2014) (Figures 1A and S1A; Table S1). Flow cytometry was used to identify and sort single FTE cells (EpCAM^+ , CD45^-), leukocytes (EpCAM^- , CD45^+), and stromal cells (EpCAM^- , CD45^-) before sequencing. To overcome the confounding batch effects and patient-specific

- (B) Diagram shows the clustering approach. Three steps are used to overcome the confounding batch effects and inter-patient variability.
- (C) Uniform manifold approximation and projection (UMAP) shows the dimensionality reduction of ~3,800 single-cell transcriptomes from fallopian tube cells. The cells are colored by their patient sources and annotated with cluster names.
- (D) Principal-component (PC) analysis plot shows the cells under different conditions (fresh, overnight cultured, and LT cultured) colored by the imputed pseudotime values.
- (E) Violin plot shows the distribution of pseudotime under different conditions. The y axis corresponds to the predicted pseudotime and the x axis the conditions. Pseudotime analysis is defined as the quantification of pseudo-temporal ordering by projecting high-dimensional data to one dimension.
- See also Figure S1; Tables S1, S2, and S3.



(legend on next page)

variability in clinical samples, we used differential-expression-based clustering (Figure 1B and the STAR Methods). Using this approach, we successfully differentiated between epithelial and non-epithelial cells (Figure 1C).

Cryopreservation and cell culture are common techniques to maintain live primary cells, especially for rare clinical samples. To test the effect of the cell maintenance approach, we compared the single-cell transcriptomes between fresh and cultured FTE cells. We observed striking effects induced by culture conditions on the transcriptomes of single cells (Figures S1B–S1H; Tables S2 and S3). For example, overnight culture induced the expression of genes that are rarely expressed in FTE cells, such as *CD44* (\log_2 fold-change [$\log\text{-FC}$] = 3.8) (Paik et al., 2012) and reduced the expression of key markers of secretory cells, such as estrogen receptor alpha (*ESR1*) and oviductal glycoprotein 1 (*OVGP1*) (Figure S1F; Table S3) (Cerny et al., 2016; Wu et al., 2016). Cilium organization was also downregulated in the overnight-cultured ciliated cells. A study showed that the Wnt pathway is essential in regulating the homeostasis of FTE cells (Kessler et al., 2015). Our analysis showed that the expression level of *LGR5* decreased in secretory cells after overnight culture, in line with the previous finding that *LGR5* was not expressed in FTE organoids (Kessler et al., 2015) (Figures S1D and S1E). The expression of *RSPO1*, another key player in the Wnt signaling, was reduced after culture, suggesting that the Wnt signaling was affected by the culture condition. Moreover, pseudotime analysis (Campbell and Yau, 2018) across three conditions revealed that the transcriptomes of long-term (LT) cultured cells were more similar to the fresh cells compared with the overnight-cultured group (Figures 1D and 1E). For example, the fatty acid metabolic process was transiently downregulated after overnight culture and then upregulated in the LT group, while the RNA processing pathway was upregulated temporarily (Figures S1B, S1C, and S1G). This suggests that including the overnight-cultured cells in subsequent analysis may introduce significant biases that would preclude meaningful conclusions. Similarly, although the LT group resembled the fresh group of cells, they showed a unique split into two subgroups and perturbed expression of Stathmin (*STMN1*) as well as cell-cycle genes that probably represented an artifact of LT culture (Figures 1D, 1E, S1B, and S1H). Therefore, overnight and LT culture likely introduced nontrivial alteration in gene expression. To avoid these substantial effects from preservation methods, we focused our downstream analysis on fresh cells only. Although fresh cells were affected by early response genes

(e.g., *FOS* and *JUN*), our analysis indicates that such affected cells can be detected by unsupervised clustering.

A Cell Census of Human Fallopian Tubes in Cancer Patients

We partitioned the fresh FTE cells (2,132) into two previously established subtypes: secretory (1,986) and ciliated cells (146) (Figure 2A). Secretory cells were characterized by the expression of *PAX8* and *KRT7* (Figures S2A and S2B) as well as a large number of newly identified markers of secretory cells (Table S4). The ciliated population was represented by the strong expression of *FOXJ1* and members of the coiled-coil domain containing protein family, such as *CCDC17* and *CCDC78* (Figures S2C and S2D). This protein family is essential for cilia functioning (Klos Dehring et al., 2013). We also identified a list of previously unrecognized markers of fallopian tube ciliated cells (Table S4), such as the calcium binding protein Calcyphosin (*CAPS*), which were enriched in the cilium-related pathways (Figures S2B, S2C, S2E, and S2F) (Wang et al., 2002).

We next identified secretory cell subtypes based on their transcriptomes. For this, only 1,410 fresh secretory cells with strong expression of *KRT7* and *EPCAM* and no expression of *CCDC17* or *PTPRC* (also known as *CD45*) were included in the analysis. In addition, to avoid including potential contaminating cancer cells, we excluded the cells that had detectable copy number variants or apparent loss-of-heterozygosity (Figure S2G) (Fan et al., 2018). We applied the aforementioned differential expression-based clustering method and identified nine clusters with distinct transcriptional profiles within the secretory cell population (Figures 2B and 2C; Table S5). Except for a patient-specific cluster (C8) that was enriched in inflammatory markers, all other clusters contained cells from multiple patients (Figures 2B, S2H, and S2I). Cluster 8 was, therefore, not considered for further analysis. Three out of nine clusters (C1, C2, and C5) had no particular distinguishing features, and clusters C1 and C2 had low library complexity (Figure S2J). This result suggests that these three clusters probably represented a quiescent population due to cell senescence or loss of hormonal influence (Crow et al., 1994; van Deursen, 2014). These clusters were, therefore, excluded from further analysis. Cluster C6 had evidence of cell stress as shown by the high expression levels of early response genes, such as *FOS* and *JUN* (Honkanиеми et al., 1992). This cluster was excluded from further analysis because such stress response is probably the result of sample preparation before sequencing.

Figure 2. Single-Cell RNA-Seq of FTE Cells Identifies Four Secretory Subtypes

- (A) UMAP plots of fresh FTE cells show the difference between secretory and ciliated cells. Each dot denotes one cell. The cells are colored by donors and the side of tubes (left panel), the expression levels of an epithelial marker, *EPCAM* (middle panel), and a secretory marker, *KRT7* (right).
- (B) t-Distributed stochastic neighbor embedding (t-SNE) plot shows the clusters within fresh secretory cells. Each point represents one cell that is colored by its cluster/subtype as shown in the legend. The quiescent population and the stress population are not shown in this plot.
- (C) Heatmap shows the clustering result of fresh secretory cells and the expression levels of their top marker genes. Each column represents a single cell and each row represents a top marker gene. Colors indicate the expression levels as shown in the scale bar. The color bar at the bottom denotes the donors.
- (D) IHC staining confirms the existence of the cell-cycle cluster (arrows) by its marker *STMN1* (Stathmin) in a human fallopian tube section.
- (E) Violin plots show the expression of nine representative marker genes of the cell-cycle cluster (C9) that are respectively related to the cell cycle, DNA repair, or chromatin remodeling pathways. Each dot is one cell.
- (F) Violin plots show the expression of six representative marker genes of the *KRT17* cluster (C4) that, respectively, belong to MHCII, cytokeratins, or ALDHs.
- (G and H) IF staining using the antibodies against *KRT17* with *PAX8* (G) or *HLA-DR* (H) in the human fallopian tube sections. Scale bars, 50 μm .
- (I) IF double staining of *KRT17* and an epithelial marker E-cadherin in the organoid derived from human FTE. Scale bars, 50 μm .

See also Figure S2; Tables S4, S5, and S6.

Cluster C9, which we termed cell-cycle cluster, comprised ~1.6% of fresh FTESCs and was enriched in three pathways, namely cell cycle (e.g., *MCM2-7*, *MKI67*, *TK1*, and *STMN1*), DNA repair (e.g., *FANCD2*, *FANCI*, and *MSH2*) and chromatin remodeling (e.g., *HMGGB2* and *SMC1A*) (Figures 2C–2E). *MKI67* (also known as Ki-67) is a well-known marker for proliferation in FTE and other cells (Kuhn et al., 2012). The two Fanconi anemia proteins, *FANCD2* and *FANCI*, can form a heterodimer that is essential for DNA repair and can interact with *MCM2-7* (Nalepa and Clapp, 2018). The relatively low percentage of cycling cells is consistent with the ages and postmenopausal status of patients from whom the cells were obtained. Cluster C3, termed the differentiated subtype, had significantly increased representation from genes involved in RNA synthesis and transport pathways (e.g., *PTBP1*, *ZNF259*, and *PRPF38A*). It also shared several markers with cluster C9. This may represent a transient differentiating cell population following cell division.

Cluster C4, termed the KRT17 subtype, was characterized by the upregulated expression of major histocompatibility complex (MHC) class II genes (e.g., *HLA-DQA1*, *HLA-DPA1*, and *HLA-DPB1*), cytokeratins (*KRT17* and *KRT23*), aldehyde dehydrogenases (ALDHs, e.g., *ALDH1A1* and *ALDH3B2*), and *CDKN1A* (also called p21) (Figures 2C and 2F; Table S6). This subpopulation of secretory FTE cells that were enriched in MHC class II expression has not been characterized previously (Comer et al., 1998). Importantly, we validated this KRT17-positive cluster in human FTE and organoids derived from human FTE cells, suggesting that it represents a robust group of cells with potentially important biological functions (Figures 2G–2I).

C7 showed high expression of a regulator of G protein signaling (*RGS16*) and genes enriched in the extracellular matrix pathway (false discovery rate [FDR] = 1.80×10^{-17}), such as *TIMP3* and *SPARC* (Figures 2C, S2K, and S2L; Table S6). We assumed that this cell type is generated by partial epithelial-mesenchymal transition (EMT) (Nieto et al., 2016), which can be induced by the chronic exposure to oxidative stress (Mahalingaiah et al., 2015) and contribute to cancer development (Hanahan and Weinberg, 2011). Hence, we termed C7 the EMT subtype. To verify that this cell type is not a contamination from FT stromal cells, we checked the expression level of epithelial markers, doublet likelihood, and the expression of stromal markers. This EMT population strongly expressed *KRT7* and *EPCAM* (Figure S2L) as expected following the aforementioned filtering. It also passed our filter for doublets; 40 (100%) out of 40 cells in the EMT cluster were detected as singlets (McGinnis et al., 2019). Moreover, two stromal markers (*COL1A2* and *COL3A1*) were specifically expressed in stromal cells but not in the EMT population (Figure S2M). These results demonstrate that this population is not contaminated by mesenchymal cells.

To exclude the potential paracrine effect of cancer cells on non-cancer FTE cells, we validated the existence of the four secretory subtypes in the FTE cells obtained from donors with benign (non-cancer) conditions. We first analyzed 1,857 single-cell transcriptomes of fallopian tubes from five patients with benign conditions (Figures 3A, S3A, and S3B; Table S1). Next we integrated the fresh secretory cells from the patients with benign conditions with the annotated ones from cancer patients by computing batch-correcting anchors (Stuart et al., 2019). Clustering of the integrated data illustrated that the four secre-

tory subtypes also existed in the FTE of non-cancer donors (Figures 3B–3D). Further validation using immunofluorescence (IF) and immunohistochemistry (IHC) in FT samples from benign donors confirmed the above results (Figures S3C–S3E). Overall, these results demonstrate that the secretory subtypes were not artifacts caused by the influence of nearby cancer cells or by systemic effects of cancer burden.

In addition to the four secretory cell types, we discovered a rare intermediate cell type that was characterized by the expression of the secretory cell marker *KRT7* and high expression of the ciliated cell marker *CAPS* (Figures 4A, 4B, S4A, and S4B), while other *KRT7*-positive secretory cells were *CAPS*-negative. *PAX8* was expressed in a subset of this intermediate population, possibly due to its moderate expression level leading to a higher dropout rate. This population was validated in human FTE tissue sections (Figures 4A and 4B). In addition, this subtype was enriched in overnight-cultured cells and recapitulated in the organoid culture derived from human FTE tissues (Figure 4C). This intermediate population most probably represents an intermediate state between secretory and ciliated cells, which accords with the previously assumed transition from secretory to ciliated cells (Ghosh et al., 2017; Hellner et al., 2016).

Because we observed that chemokines and MHC genes were frequently expressed in FTE cells (Figure 2B), we examined fallopian tube sections for the expression of lymphocyte markers and identified a basal *CD45*⁺ *EpCAM*⁺ cell population. This population was also positive for *CD3*, *CD44*, *CD69*, and *CD103* (Figures 4D–4H), suggesting that these basal cells are tissue-resident memory T lymphocytes (TRMs) (Topham and Reilly, 2018) and that the FTE is not immunologically inert. This is in line with previous studies (Ardighieri et al., 2014; Peters, 1986). Nevertheless, we demonstrate that this basal cell population was also *EpCAM*⁺, which is surprising because TRMs have not been reported to be positive for epithelial markers in human.

Revealing the Cell State Composition of HGSOCs Using Deconvolution

Given the link between types of cell-of-origin and tumor types that has been reported in other cancers, such as breast cancer (Bertucci et al., 2012; Dai et al., 2015) and the known shared key markers between HGSOC and FTE cells (e.g., *PAX8*, *WT1*, and *ESR1*) (Ince et al., 2015; Perets et al., 2013), we hypothesized that HGSOC cell states are linked to FTE cell subtypes. Based on our profiling of single-cell transcriptomes, we computed a reference matrix with cell-type-derived transcriptomic signatures from the five major FTE cellular subtypes (cell cycle, EMT, differentiated, KRT17 cluster, and ciliated) as described previously (Baron et al., 2016) (Figure 5A; Table S7). The resulting reference matrix was then used in the deconvolution analysis (Newman et al., 2015) of the bulk HGSOC RNA-seq data from The Cancer Genome Atlas (TCGA) (Bell et al., 2011) and the microarray data from the Australian Ovarian Cancer Study (AOCS) (Tothill et al., 2008) to compute the fractions of five cell states within each tumor. The decomposition of bulk tumor samples revealed intra-tumor transcriptomic heterogeneity according to the proportions of cell subtypes in both datasets (Figures 5B–5E). Remarkably, the ciliated tumor subtype was highly enriched in the low-grade tumors (grade 1) compared with the high-grade ones (grades 2–3) in the AOCS dataset

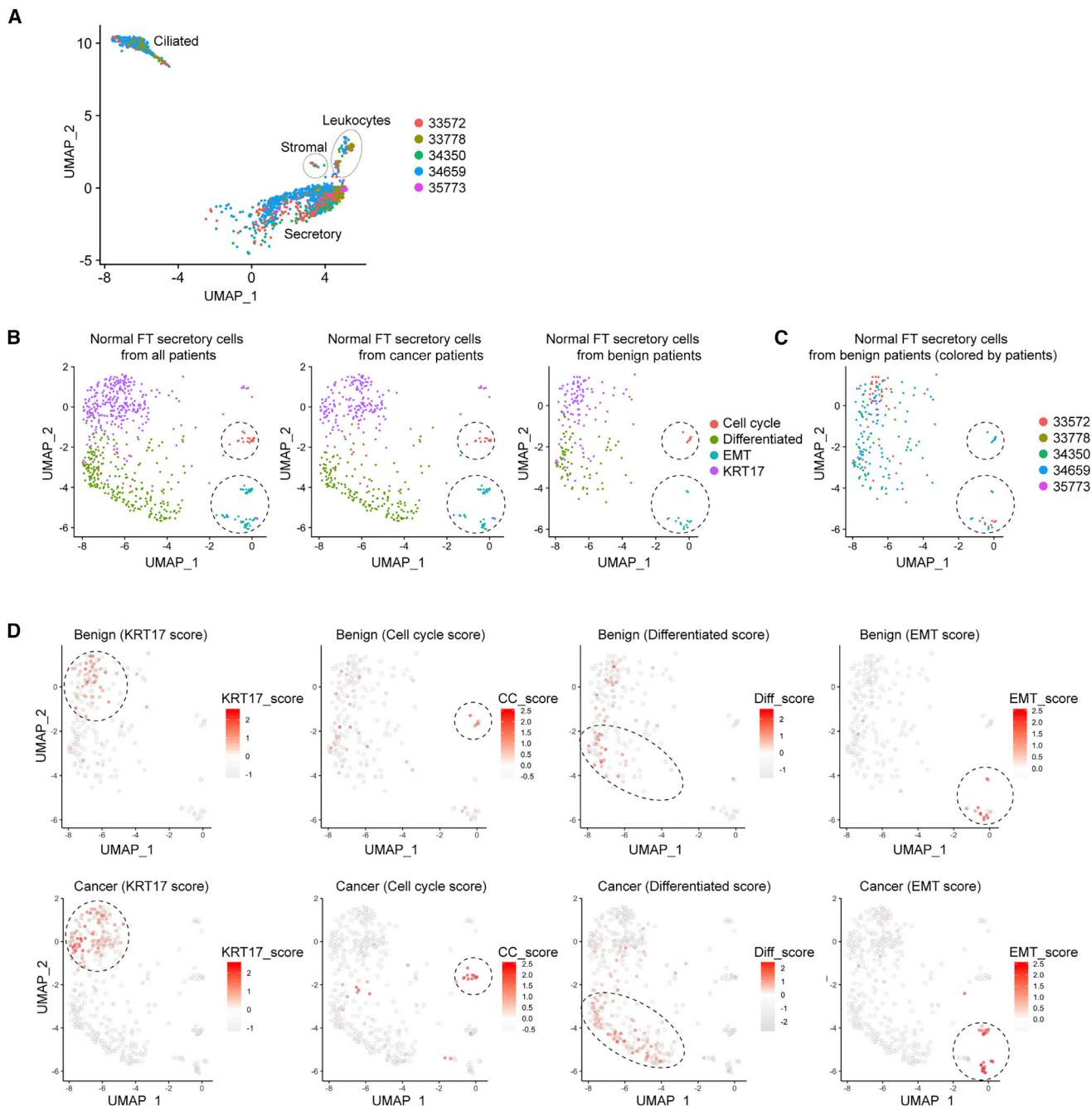


Figure 3. Validation of the Secretory Subtypes in the FTE of Benign Donors by Using scRNA-Seq

- (A) UMAP shows the populations in the FT samples from five patients with benign conditions. Each dot is a cell colored by its donor.
- (B) UMAP plots show the populations in the FT samples from cancer patients ($n = 5$) and patients with benign conditions ($n = 5$). The left, middle, and right subpanels contain the cells from all ten patients, five cancer patients, and five patients with benign conditions, respectively.
- (C) UMAP plot shows the populations in the FT samples from five patients with benign conditions. Each dot represents a secretory cell from a benign patient. The dots are colored by their donors as shown in the legend.
- (D) Scatterplots show the transcriptomic characteristics of each subtype in benign and cancer patients. Cells (dots) are colored by the score of each transcriptomic signature (subtitle). The score of a transcriptomic signature was computed by the scaled and centered sum of expression levels of the marker genes in each transcriptomic signature. The scores correspond to the expression of marker genes of each cluster. The transcriptomic signatures are listed in Table S7.

See also Figure S3.

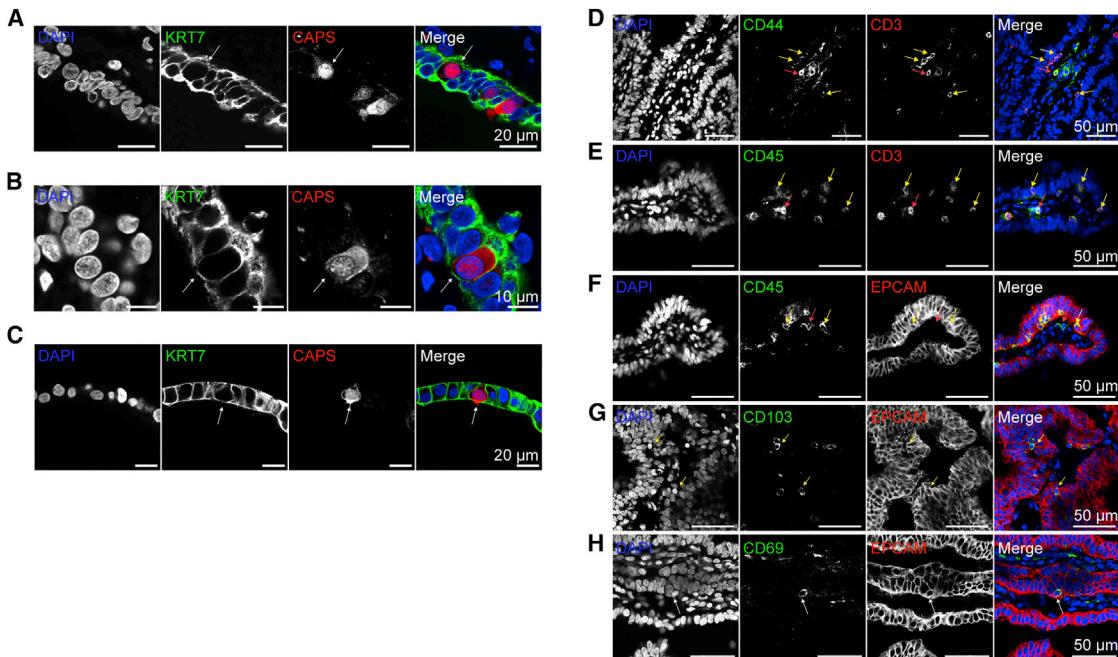


Figure 4. Non-traditional Cell Subtypes in the Fallopian Tube Epithelial Layer

(A and B) IF staining for KRT7 and CAPS in human FTE samples (A and B). Arrows indicate the double-positive intermediate cell population. (C) IF staining for KRT7 and CAPS in the cultured human fallopian tube organoid. Arrow indicates the KRT7⁺ CAPS⁺ intermediate cell population. (D–F) IF staining of CD44 and CD3 (D), CD45 and CD3 (E), as well as CD45 and EpCAM (F) in human FTE. Yellow arrows indicate the intra-epithelial CD44⁺ CD3⁺ cells that are also CD45⁺ and EpCAM⁺. Red arrows indicate the extra-epithelial CD44⁺ CD3⁺ CD45[−] EpCAM[−] cells in the stromal region. (G and H) IF staining for EpCAM and two markers of tissue-resident memory T lymphocytes, CD103 (G), and CD69 (H), in the human FTE. See also Figure S4.

and an additional dataset that also comprised both high- and low-grade tumors (Yoshihara et al., 2010) ($p = 1.3e-10$, one-sided Wilcox test, Figures 5F and S5). This strongly suggests that grades of SOC are associated with their ability to differentiate into cells that molecularly resemble FTE ciliated cells. In contrast, the TCGA dataset, which only included HGSOCS, had no tumors that were enriched in the ciliated subtype (Figure 5B). Therefore, the enrichment in these ciliated markers is most probably a distinguishing molecular feature of low-grade tumors.

Most notably, we identified a class of EMT-enriched tumors in multiple datasets. These tumors were enriched in the genes previously linked to the “mesenchymal” HGSOCS subtype. We found that the marker genes of these tumors were enriched in the extracellular matrix, focal adhesion, and PI3K-AKT signaling pathways (FDR < 0.0002, by DAVID, Table S8), that are critical for tumor cell survival (Fresno Vara et al., 2004; McLean et al., 2005). Furthermore, three key EMT transcription factors, TWIST1, TWIST2, and SNAI2 (Ansieau et al., 2008; Kang and Massagué, 2004; Yang et al., 2004), were upregulated in the EMT-high tumors (Figure 5G), while the miRNA-200 family (miR-200a, miR-200b, miR-200c, miR-141, and miR-429) that represses EMT (Gregory et al., 2008; Nieto et al., 2016) was downregulated in EMT-high tumors (FDR < 0.01, log-FC < -0.5, Figure 5H), suggesting that EMT may be the underlying mechanism of enrichment of mesenchymal cancer cells in this tumor subtype. We also found that miRNA-483 and miRNA-214 were significantly upregulated in EMT-high tumors, while

miRNA-513c, miRNA-509, and miRNA-514 were downregulated (Figure 5H). Although previous studies suggested that miRNA-483 and miRNA-214 play an important role in cancer progression (Chandrasekaran et al., 2016; Liu et al., 2013), their regulatory role in the EMT process has not been evaluated.

EMT-High Subtype Is Robustly Correlated with Poor Prognosis

We next tested whether any of the five tumor subtype scores from the deconvolution analysis was associated with survival. We found that the EMT score was significantly associated with poor overall survival and that this association was independent from the effect of age, stage, or residual disease ($p < 0.05$, by Cox proportional hazard model, Table 1). The robustness of this association was confirmed by permutation testing ($n = 500$) leaving out 10% of the samples each time (empirical p values = 0.012 [TCGA] and 0 [AOCS], by permutation test). The mesenchymal subtype was reported previously to be associated with poor prognosis (Konecny et al., 2014; Tan et al., 2013), but the reproducibility of this observation was inconsistent probably because of the difficulty in defining this group of tumors. We first compared the prognosis of EMT-high tumors with the previously defined mesenchymal HGSOCS subtype in the TCGA data. The analysis showed that the EMT score had higher sensitivity of identifying poor-prognostic patients (Figure S6A). Importantly, among the 238 cancers that were labeled as non-mesenchymal in TCGA we identified 87 tumors that were EMT-high. These tumors carried worse prognosis when compared

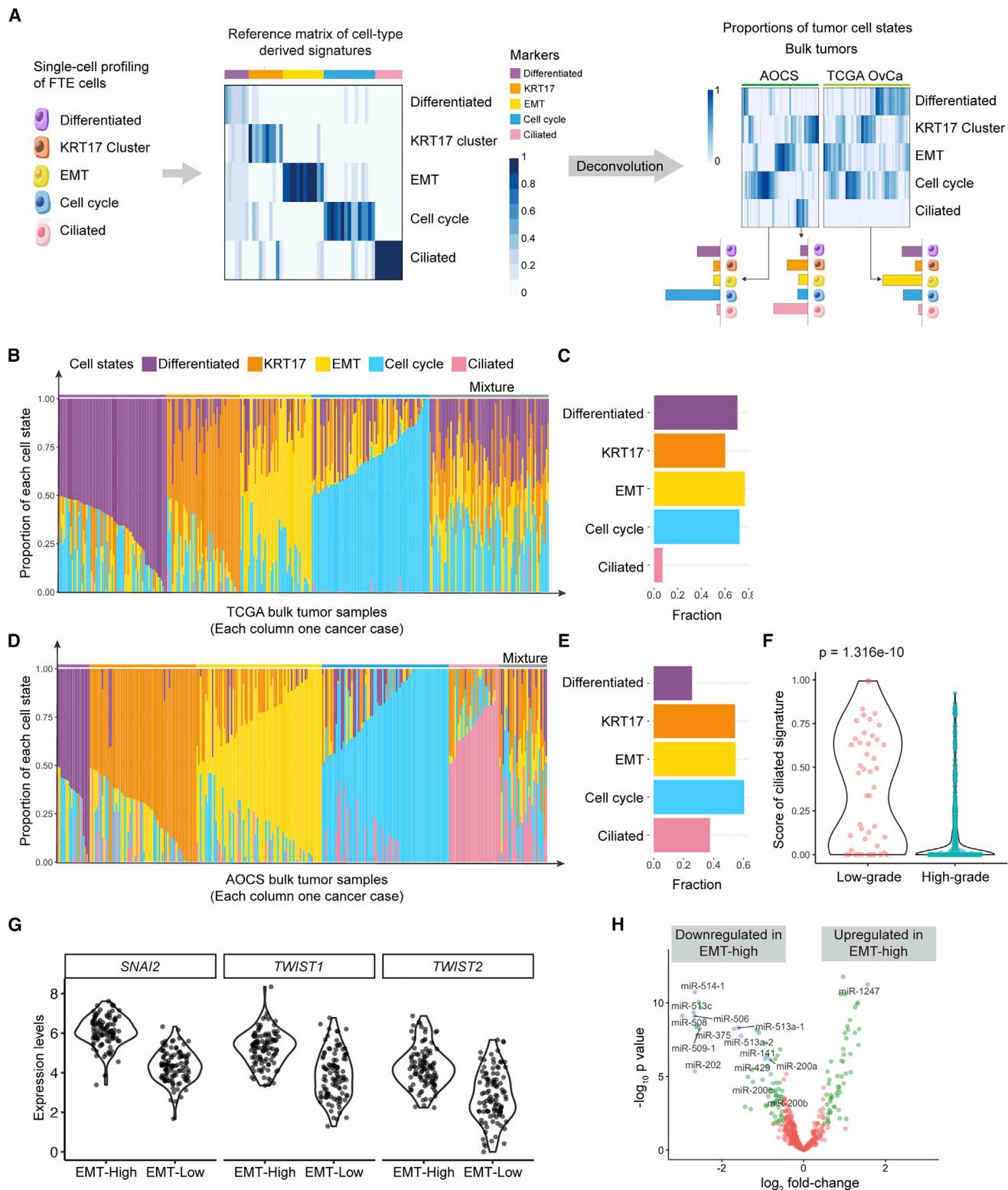


Figure 5. The Repertoire of Phenotypic Heterogeneity of Serous Ovarian Cancer Revealed by Deconvolution Analysis

(A) A reference matrix was calculated from our scRNA-seq data of FTE cells, where columns correspond to 5 cell subtypes in FTE as indicated. The reference matrix heatmap depicts the magnitude of the expression levels of the 52 marker genes (rows) across 5 cell subtypes, including 7 differentiated markers, 10 KRT17 cluster markers, 12 EMT markers, 15 cell-cycle markers, and 8 ciliated markers. The heatmaps in the right panel depict the proportions of cell states (rows) across

(legend continued on next page)

Table 1. EMT-High Serous Tumors Are Robustly Correlated with Poor Prognosis

Dataset	Samples	Hazard Ratio	p	95% CI	Citation
E.MTAB.386	128	3.11	0.018	1.21–7.95	Bentink et al., 2012
GSE49997	170	2.83	0.022	1.16–6.88	Pils et al., 2012
GSE13876	144	2.96	0.011	1.28–6.83	Crijns et al., 2009
GSE26712	185	2.03	0.031	1.07–3.89	Bonomé et al., 2008
GSE26193	79	2.58	0.041	1.04–6.42	Mateescu et al., 2011
GSE51088	113	2.09	0.022	1.11–3.92	Karlan et al., 2014
GSE32062.GPL6480	260	1.89	0.055	0.99–3.60	Yoshihara et al., 2012
AOCS	253	2.11	0.008	1.21–3.67	Tothill et al., 2008
TCGA	305	2.12	0.011	1.18–3.81	Bell et al., 2011

See also Table S9.

with the non-mesenchymal and EMT-low group of cancer. Furthermore, by using EMT scores from deconvolution, we reached a robust classification with consistently significant correlation with poor survival ($p < 0.05$) in eight independent datasets, including TCGA dataset, AOCS dataset, and six additional microarray datasets ($n > 100$ patients in each set) from the CuratedOvarianData database (Ganzfried et al., 2013) (Tables 1 and S9). When we combined and dichotomized all the samples ($n = 1,626$), the EMT-high tumors had significantly worse prognosis ($p = 5.92e-10$; hazard ratio = 1.5; 95% confidence interval [CI], 1.3–1.7 by log rank test, Figure 6A). This demonstrates that deconvolution analysis using the identified panel of genes is strongly predictive of prognosis in SOC.

To test whether the EMT-high tumor subtype was merely a reflection of stromal cell impurities in tumor samples, we performed RNA sequencing on 36 laser capture microdissected (LCM) tumor samples collected from 15 patients with HGSOC (termed the Oxford Ovarian Cancer-Predict Chemotherapy Response [OXO-PCR] cohort) and classified tumors based on the deconvolution analysis (Figure 6B). We next compared the expression of genes in the EMT signature between LCM tumor and stroma samples. As expected, the expression levels of *PAX8* and *EPCAM* were significantly higher in both EMT-high and EMT-low tumor samples compared with stroma in which these markers showed almost no expression (Figure S6B). In contrast, the EMT markers (*SPARC*, *TIMP3*, *DCN*, and *SFRP4*) were highly expressed in EMT-high tumors confirming that

EMT-high tumors truly expressed these genes (Figure 6C), further confirming EMT-high as a genuine tumor subtype. *SPARC*, one of the 12 genes that constitute the EMT signature, was described previously in the mesenchymal subtype of HGSOC (Tothill et al., 2008), while certain other markers were reported to be related to EMT in ovarian cancer or other cancers, such as *SFRP4* (Ford et al., 2013), *TIMP3* (Anastassiou et al., 2011), *MYH11* (Li and Yang, 2016), and *EFEMP1* (Yin et al., 2016). Nevertheless, the link between this tumor subtype and a particular FTE cellular subtype was previously unrevealed. To further confirm that the EMT-high marker *SPARC* is strongly expressed in cancer cells, we performed immunofluorescent staining on pre-chemotherapy frozen tumor sections, where we indicated cancer cells by co-staining with an antibody that recognized pan-cytokeratin. Our results revealed the co-existence of at least two populations of cancer cells, a pan-cytokeratin-high, *SPARC*-negative population and a pan-cytokeratin-low and *SPARC*-positive population within the same tumor (Figures 6D and S6C; Video S1). In addition, pan-cytokeratin-negative and *SPARC*-high cancer cells could also be identified by their abnormal nuclei (Figures 6E and S6D), which were distinct from the *SPARC*-high fibroblasts cells (Tan et al., 1999). We also confirmed the expression of *SPARC* in p53-positive tumor cells by performing IHC on p53-expressing HGSOC samples (Figures S6E and S6F).

Similarly, this *SPARC*-positive, pan-cytokeratin-negative, and mesenchymal-like subpopulation was observed in the

bulk tumor samples (columns) in AOCS and TCGA datasets. Beneath these heatmaps, three cartoon bar plots that correspond to three samples show a schematic composition of these samples.

(B) Stacked bar plot shows the deconvolution result of 308 tumors from the TCGA ovarian carcinoma study. Colors of the bars denote 5 cell states as shown in the legend. The y axis represents the proportion of each state in a given bulk tumor sample. In the x axis, each column represents one tumor case. The annotation bar above denotes the subtypes of bulk tumors that are defined by the dominant cell state within each tumor, where gray represents the mixture of multiple cell states. (C) Bar plot shows the prevalence of 5 cell states in the TCGA dataset. The x axis represents the proportion of tumor samples in TCGA that harbor the given cell state (y axis).

(D) Stacked bar plot shows the deconvolution result of 285 tumors from the AOCS dataset as in (B).

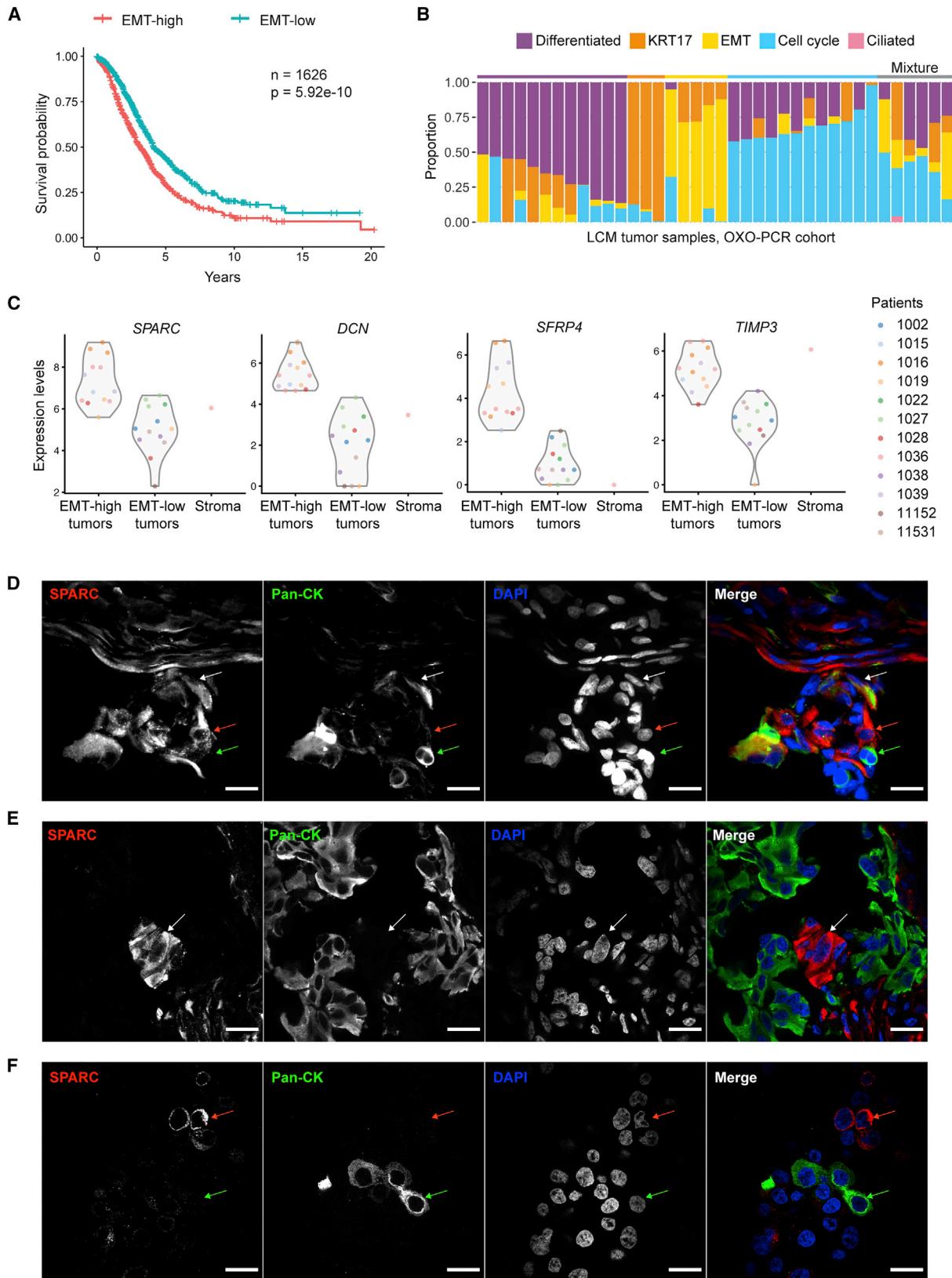
(E) Bar plot shows the prevalence of 5 cell states in the AOCS dataset.

(F) Violin plot shows the ciliated scores in low-grade SOC compared with HGSOC. We combined samples from both AOCS (20 low-grade, 260 high-grade) (Tothill et al., 2008) and GSE17260 (26 low-grade, 84 high-grade) (Yoshihara et al., 2010). Each dot is one sample, colored according to grade. $p = 1.3e-10$, by one-sided Wilcoxon rank-sum test.

(G) Violin plots show expression of *TWIST1*, *TWIST2*, and *SNAI2* in the EMT-high tumors compared with the EMT-low ones in the TCGA dataset (log-FC > 1.8, FDR < 2×10^{-14} , by limma voom).

(H) Volcano plot shows the miRNAs that are differentially expressed between EMT-high and EMT-low tumors in the TCGA dataset. The green and blue dots correspond to the miRNAs that are significantly differentially expressed (log-FC > 0.5, FDR < 0.05).

See also Figure S5 and Table S7.



(legend on next page)

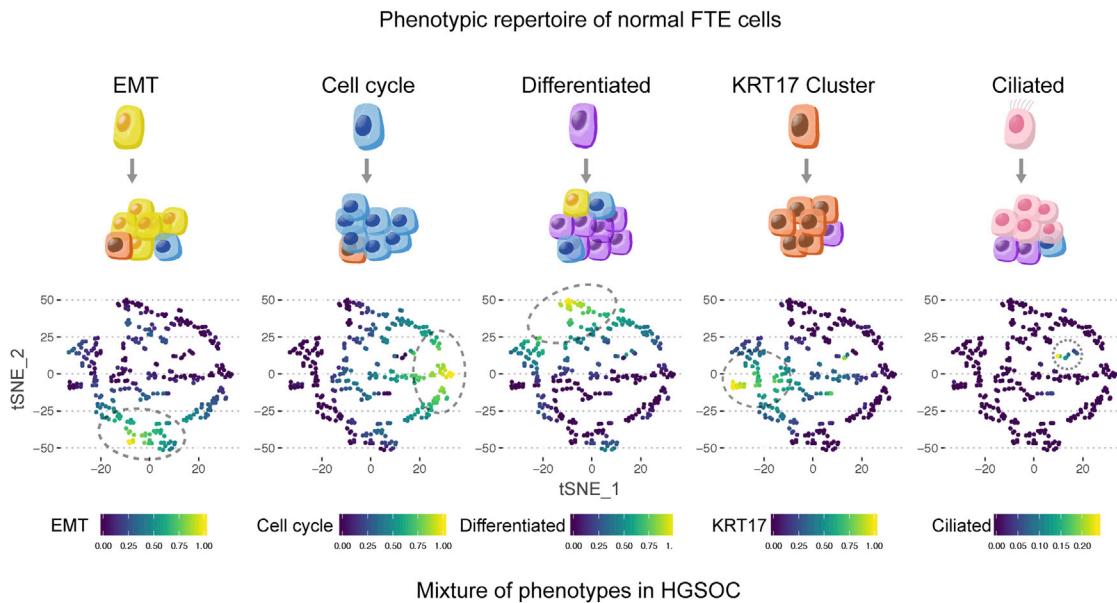


Figure 7. HGSOCs Are Composed of a Mixture of Cancer Cell States Inherited from Its Cell-of-Origin in the FTE

The diagram proposes a model that HGSOCs are assembled from multiple cancer cell states that echo the cellular phenotypes of its tissue-of-origin. The upper panel exhibits five cellular subtypes of FTE. In the lower panel, five t-SNE plots demonstrate the cell state hubs within the deconvolution result of TCGA data. Each dot is an HGSOC bulk sample. The color scale denotes the proportion of cancer cell states. Gray dashed circles highlight the hubs representing the tumor subtypes of SOC.

KURAMOCHI cell line, which is known to faithfully represent HGSOC (Domcke et al., 2013). KURAMOCHI cells exhibited two phenotypes, a mesenchymal-like phenotype that was SPARC positive and pan-cytokeratin negative and an epithelial phenotype that was SPARC negative and pan-cytokeratin positive (Figures 6F, S6G, and S6H), suggesting that this cell line could be a potential *in vitro* model to study the mesenchymal subtype of HGSOC. Importantly, SPARC-positive cells expressed p53 (Figure S6I) providing further confirmation of the existence of EMT-high cancer cells. Moreover, TP53 sequencing of this cell line showed a single peak for the previously described mutation (c.841G>T) in this cell line (Anglesio et al., 2013), validating that the phenotypic heterogeneity was not caused by contamination with another cell line.

DISCUSSION

In this study, we performed deep single-cell RNA-seq on thousands of nonmalignant epithelial cells from the FTE and uncovered a strong link between diverse secretory cell populations and HGSOC subtypes. This elucidates the repertoire of phenotypic heterogeneity in cancer cells that are inherited from the cell-of-origin. We propose a model whereby individual HGSOCs are composed of a mixture of cancer cell states that we revealed by single-cell sequencing of the putative cell-of-origin (Figure 7). Unraveling the trajectory of differentiation from one cancer cell state to another will be needed in future studies to understand the mechanistic basis of HGSOC cell plasticity.

Figure 6. EMT-High Tumors Comprise Mesenchymal-Like Cancer Cells

- (A) Kaplan-Meier curves show the effect of EMT scores on survival. The cases of SOC ($n = 1,626$) are combined from nine datasets (Table 1). Patients are dichotomized into EMT-high and EMT-low by the median of EMT scores in each dataset. Each short vertical line indicates a censoring event. $p = 5.92 \times 10^{-10}$, hazard ratio = 1.5 (95% CI, 1.3–1.7), by log rank test.
- (B) Stacked bar plot visualizes the deconvolution result of 36 LCM tumor samples collected from 15 patients with HGSOC in the OXO-PCR cohort. The y axis denotes the proportion (0–1) of the 5 cell states (colors) across LCM tumor samples (rows). The annotation bar above the bar plot corresponds to the tumor subtypes that are defined by the dominant cell state within each tumor, where gray represents the mixture of multiple cell states.
- (C) Violin plots show the expression levels of four EMT markers in the EMT-high LCM tumor samples compared with the EMT-low ones with one LCM stromal sample as control ($p < 0.001$, one-sided Wilcoxon test). Each dot is a sample colored by patients.
- (D) IF staining in the pre-chemotherapy HGSOC tumor frozen section, using an EMT marker, SPARC, and an epithelial marker, pan-cytokeratin (Pan-CK). White arrows indicate cells that are double-positive for SPARC and Pan-CK; red arrows indicate cells that are SPARC⁺ only; yellow arrows indicate cells that are Pan-CK⁺ only. Scale bars, 20 μ m.
- (E) IF staining for SPARC and Pan-CK in the pre-chemotherapy HGSOC tumor frozen section. Arrows show a tumor cell that is SPARC⁺ and Pan-CK⁻. Scale bars, 20 μ m.
- (F) IF staining for SPARC and Pan-CK in the KURAMOCHI cell line. Red arrows indicate cells that are SPARC⁺ and Pan-CK⁻; green arrows indicate cells that are SPARC⁻ Pan-CK⁺. Scale bars, 20 μ m.

See also Figure S6; Table S8; Video S1.

Within the FTE, our analyses recognized a KRT17-positive population with high expression of ALDHs and MHCII. ALDH was reported as the marker of mammary stem cells (Biton et al., 2018; Ginestier et al., 2007). Moreover, a subpopulation of intestinal stem cells was recently reported to express MHCII. The latter was found to regulate the stem cell fate through the interaction between T helper cells and stem cells (Biton et al., 2018; Ginestier et al., 2007). Such evidence implies that the KRT17 cluster that we identified could include a progenitor population. This opens a door for studying self-maintenance and differentiation of epithelial cells in fallopian tubes.

Phenotypic diversity exists in HGSOC at the transcriptomic level (Bell et al., 2011; Tothill et al., 2008). Based on the single-cell profiling of FTE cells, our deconvolution analysis revealed the association between the FTE cellular subtypes with ovarian cancer cell states within individual tumors. The correspondence between tumors and normal tissues can be explained by at least two hypotheses that are not necessarily mutually exclusive. The first one is that each tumor subtype originates from an individual cell type of origin (Ince et al., 2007; Visvader, 2011). The second hypothesis is that multiple tumor subtypes come from one single cellular subtype, sometimes referred to as the cancer stem cell, while the plasticity of cancer cells mimics that of the normal cell-of-origin (Gupta et al., 2011; Visvader, 2011). Both concepts are compatible with our findings. Lineage tracing will be needed to investigate the underlying mechanism of the correspondence between tumors and the tissue-of-origin. Although the FTE plays as a strong candidate for the cell-of-origin of HGSOC, there is a possibility that the ovarian surface epithelium gives rise to a proportion of HGSOC cases as supported by the mouse model (Zhang et al., 2019b). The association between ovarian surface epithelium and HGSOC has not been addressed by our work.

Despite the fact that the mesenchymal-like tumor subtype was previously reported as a distinct subtype in HGSOC (Tothill et al., 2008), it has been controversial whether the expression of EMT markers originated from tumor cells, infiltrating stromal cells or both (Schwede et al., 2018; Zhang et al., 2019a). Our results demonstrate that the expression of mesenchymal-related genes, such as *SPARC*, occurs in tumor cells. This further supports the existence of the EMT process in human tumors, which may enhance their invasion ability and drug resistance (Schmidt et al., 2015; Singh and Settleman, 2010). This notion is consistent with our finding that the EMT-high HGSOC subtype is related to poor survival. However, we cannot exclude the possibility that the expression of EMT genes in bulk tumor analyses may arise from stromal infiltration or cancer-associated fibroblasts (CAFs). Since CAFs can potentially promote cancer progression (Chen et al., 2018; Kalluri, 2016; Schwede et al., 2018; Eckert et al., 2019), it will also be important to scrutinize the role of CAFs in HGSOC and the interaction between CAFs and ovarian cancer cells.

Compared with the previously developed prognostic signatures of HGSOC that were difficult to be reproduced in multiple datasets (Bell et al., 2011; Bonome et al., 2008; Chen et al., 2018; Jiang et al., 2006; Karlan et al., 2014; Konecny et al., 2014), our classifier (we term, the Oxford Classic [Oxford Classifier of Carcinoma of the Ovary]) is strongly reproducible across 8 independent datasets and is based on the expression of only 52 marker genes. Robust identification of tumor subtypes is a pre-

requisite for identifying treatment strategies that limit the ability of cancer cell to acquire a mesenchymal state (Davis et al., 2014; Ishay-Ronen et al., 2019). Given these factors, our approach possesses potential translational significance to improve the survival of HGSOC patients.

In conclusion, our single-cell profiling illuminates the cellular landscape of human FTE, the presumed tissue-of-origin of HGSOC. Our work not only expands the understanding of FTE, but also provides a benchmark dataset on FTE to study HGSOC. The combined analysis of single-cell data and tumor expression data revealed the connection between FTE cellular subtypes and HGSOC tumor subtypes and identified the EMT-high subtype with important prognostic implications.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Cell Line
 - Human Fallopian Tube and Tumor Samples
- METHOD DETAILS
 - Sample Dissociation
 - Culture, Cryopreservation and Harvest
 - Organoid Culture
 - Cell Sorting
 - cDNA Synthesis and Library Construction
 - Whole-Exome Sequencing
 - Laser Capture Microdissection
 - Bulk RNA Sequencing
 - Immunofluorescent Staining
 - Immunohistochemistry
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Preprocessing scRNA-seq Data
 - Detecting Doublets
 - Visualization of Expression Data
 - Comparing Fresh and Cultured Cells
 - Pseudotime Analysis
 - Filtering to Select Fresh Secretory Cells
 - Differentially-Expressed-Based Clustering
 - Data Integration
 - Identifying Marker Genes
 - Gene Ontology Enrichment Analysis
 - Bulk Expression Data
 - Deconvolution of Bulk Expression Data
 - Survival Analysis
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.ccr.2020.01.003>.

ACKNOWLEDGMENTS

We thank the WIMM Flow Cytometry Facility, Single Cell Facility, CBRG, and Wolfson Imaging Centre for their help in this study. We thank the Oxford

Radcliffe Biobank and the Oxford Centre for Histopathology Research for their help with obtaining tissue samples. This work was supported by Ovarian Cancer Action and Oxford Biomedical Research Centre (BRC), National Institute for Health Research. C.Y. was supported by UK Medical Research Council Research Grants (MR/L001411/1 and MR/P02646X/1), a Wellcome Trust Core Award grant number 090532/Z/09/Z, and by The Alan Turing Institute under EPSRC grant EP/N510129/1. Z.H. was supported by the NIHR Oxford BRC. D.M.P. was funded by the NIHR Oxford BRC (Molecular Diagnostics Theme/Multimodal Pathology Subtheme). Y.F. was supported by the CRUK Oxford Centre.

AUTHOR CONTRIBUTIONS

Conceptualization, Project Administration, and Supervision, A.A.A. and C.Y.; Funding Acquisition, A.A.A., C.Y., and Y.F.; Investigation, A.A.A., C.Y., Z.H., M.A., A.A., N.W., M.M., L.S.G., S.E., M.K., G.M., T.S., Z.Z., K.M., Y.Z., K.C., V.C., and T.S.-S.; Data Curation, A.A.A., C.Y., Z.H., and M.A.; Methodology and Formal Analysis and Visualization, A.A.A., C.Y., Z.H., and A.A.; Writing – Original Draft, A.A.A., C.Y., Z.H., M.A., and A.A.; Writing – Review & Editing, A.A.A., C.Y., and Z.H.; Resources (Clinical Samples), M.M., S.Damato, S.Dhar, V.R., R.G.C., H.S., D.M.-P., and S.J.

DECLARATION OF INTERESTS

A.A.A., C.Y., and Z.H. filed a patent application for the use of the 52-gene panel for predicting the prognosis in ovarian cancer.

Received: May 24, 2019

Revised: October 30, 2019

Accepted: January 9, 2020

Published: February 10, 2020

REFERENCES

- Ahmed, A.A., Etemadmoghadam, D., Temple, J., Lynch, A.G., Riad, M., Sharma, R., Stewart, C., Fereday, S., Caldas, C., DeFazio, A., et al. (2010). Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. *J. Pathol.* 221, 49–56.
- Anastassiou, D., Rumjantseva, V., Cheng, W., Huang, J., Canoll, P.D., Yamashiro, D.J., and Kandel, J.J. (2011). Human cancer cells express Slug-based epithelial-mesenchymal transition gene expression signature obtained in vivo. *BMC Cancer* 11, 529.
- Anglesio, M.S., Wiegand, K.C., Melnyk, N., Chow, C., Salamanca, C., Prentice, L.M., Senz, J., Yang, W., Spillman, M.A., Cochrane, D.R., et al. (2013). Type-specific cell line models for type-specific ovarian cancer research. *PLoS One* 8, e72162.
- Ansieau, S., Bastid, J., Doreau, A., Morel, A.-P., Bouchet, B.P., Thomas, C., Fauvet, F., Puisieux, I., Doglioni, C., Piccinin, S., et al. (2008). Induction of EMT by twist proteins as a collateral effect of tumor-promoting inactivation of premature senescence. *Cancer Cell* 14, 79–89.
- Ardighieri, L., Lonardi, S., Moratto, D., Facchetti, F., Shih, I.-M., Vermi, W., and Kurman, R.J. (2014). Characterization of the immune cell repertoire in the normal fallopian tube. *Int. J. Gynecol. Pathol.* 33, 581–591.
- Baron, M., Veres, A., Wolock, S.L., Faust, A.L., Gaujoux, R., Vetere, A., Ryu, J.H., Wagner, B.K., Shen-Orr, S.S., Klein, A.M., et al. (2016). A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell Syst.* 3, 346–360.e4.
- Bell, D., Berchuck, A., Birrer, M., Chien, J., Cramer, D.W., Dao, F., Dhir, R., DiSaia, P., Gabra, H., Glenn, P., et al. (2011). Integrated genomic analyses of ovarian carcinoma. *Nature* 474, 609–615.
- Bentink, S., Haibe-Kains, B., Risch, T., Fan, J.B., Hirsch, M.S., Holton, K., Rubio, R., April, C., Chen, J., Wickham-Garcia, E., et al. (2012). Angiogenic mRNA and microRNA gene expression signature predicts a novel subtype of serous ovarian cancer. *PLoS One* 7, e30269.
- Bertucci, F., Finetti, P., and Birnbaum, D. (2012). Basal breast cancer: a complex and deadly molecular subtype. *Curr. Mol. Med.* 12, 96–110.
- Biton, M., Haber, A.L., Rogel, N., Burgin, G., Beyaz, S., Schnell, A., Ashenberg, O., Su, C.-W., Smillie, C., Shekhar, K., et al. (2018). T helper cell cytokines modulate intestinal stem cell renewal and differentiation. *Cell* 175, 1307–1320.e22.
- Bonomé, T., Levine, D.A., Shih, J., Randonovich, M., Pise-Masison, C.A., Bogomolniy, F., Ozburun, L., Brady, J., Barrett, J.C., Boyd, J., et al. (2008). A gene signature predicting for survival in suboptimally debulked patients with ovarian cancer. *Cancer Res.* 68, 5478–5486.
- Brock, A., Chang, H., and Huang, S. (2009). Non-genetic heterogeneity—mutation-independent driving force for the somatic evolution of tumours. *Nat. Rev. Genet.* 10, 336–342.
- Butler, A., Hoffman, P., Hoffman, P., Smibert, P., Smibert, P., Papalexi, E., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36, 411–420.
- Campbell, K.R., and Yau, C. (2018). Uncovering pseudotemporal trajectories with covariates from single cell and bulk expression data. *Nat. Commun.* 9, 2442.
- Cerny, K.L., Ribeiro, R.A.C., Jeoung, M., Ko, C., and Bridges, P.J. (2016). Estrogen receptor alpha (ESR1)-dependent regulation of the mouse oviductal transcriptome. *PLoS One* 11, e0147685.
- Chandrasekaran, K.S., Sathyaranayanan, A., and Karunagaran, D. (2016). MicroRNA-214 suppresses growth, migration and invasion through a novel target, high mobility group AT-hook 1, in human cervical and colorectal cancer cells. *Br. J. Cancer* 115, 741–751.
- Chen, G.M., Kannan, L., Geistlinger, L., Kofia, V., Safikhani, Z., Gendoo, D.M.A., Parmigiani, G., Birrer, M., Haibe-Kains, B., and Waldron, L. (2018). Consensus on molecular subtypes of high-grade serous ovarian carcinoma. *Clin. Cancer Res.* 24, 5037–5047.
- Clyman, M.J. (1966). Electron microscopy of the human fallopian tube. *Fertil. Steril.* 17, 281–301.
- Comer, M.T., Leese, H.J., and Southgate, J. (1998). Induction of a differentiated ciliated cell phenotype in primary cultures of fallopian tube epithelium. *Hum. Reprod.* 13, 3114–3120.
- Crijns, A.P.G., Fehrmann, R.S.N., de Jong, S., Gerbens, F., Meersma, G.J., Klip, H.G., Hollema, H., Hofstra, R.M.W., Meerman te, G.J., de Vries, E.G.E., et al. (2009). Survival-related profile, pathways, and transcription factors in ovarian cancer. *PLoS Med.* 6, e24.
- Crow, J., Amso, N.N., Lewin, J., and Shaw, R.W. (1994). Morphology and ultrastructure of fallopian-tube epithelium at different stages of the menstrual-cycle and menopause. *Hum. Reprod.* 9, 2224–2233.
- Dai, X., Li, T., Bai, Z., Yang, Y., Liu, X., Zhan, J., and Shi, B. (2015). Breast cancer intrinsic subtype classification, clinical use and future trends. *Am. J. Cancer Res.* 5, 2929–2943.
- Davis, F.M., Stewart, T.A., Thompson, E.W., and Monteith, G.R. (2014). Targeting EMT in cancer: opportunities for pharmacological intervention. *Trends Pharmacol. Sci.* 35, 479–488.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2012). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Domcke, S., Sinha, R., Levine, D.A., Sander, C., and Schultz, N. (2013). Evaluating cell lines as tumour models by comparison of genomic profiles. *Nat. Commun.* 4, 2126.
- Ducie, J., Dao, F., Considine, M., Olvera, N., Shaw, P.A., Kurman, R.J., Shih, I.-M., Soslow, R.A., Cope, L., and Levine, D.A. (2017). Molecular analysis of high-grade serous ovarian carcinoma with and without associated serous tubal intra-epithelial carcinoma. *Nat. Commun.* 8, 990.
- Eckert, M.A., Coscia, F., Chryplewicz, A., Chang, J.W., Hernandez, K.M., Pan, S., Tienda, S.M., Nahotko, D.A., Li, G., Blaženović, I., et al. (2019). Proteomics reveals NNM1 as a master metabolic regulator of cancer-associated fibroblasts. *Nature* 474, 609–728.
- Eden, E., Lipson, D., Yogeve, S., and Yakhini, Z. (2007). Discovering motifs in ranked lists of DNA sequences. *PLoS Comput. Biol.* 3, 508–522.

- Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10, 48.
- Etemadmoghadam, D., DeFazio, A., Beroukhim, R., Mermel, C., George, J., Getz, G., Tothill, R., Okamoto, A., Raeder, M.B., Harnett, P., et al. (2009). Integrated genome-wide DNA copy number and expression analysis identifies distinct mechanisms of primary chemoresistance in ovarian carcinomas. *Clin. Cancer Res.* 15, 1417–1427.
- Fan, J., Lee, H.-O., Lee, S., Ryu, D.-E., Lee, S., Xue, C., Kim, S.J., Kim, K., Barkas, N., Park, P.J., et al. (2018). Linking transcriptional and genetic tumor heterogeneity through allele analysis of single-cell RNA-seq data. *Genome Res.* 28, 1217–1227.
- Fessler, E., and Medema, J.P. (2016). Colorectal cancer subtypes: developmental origin and microenvironmental regulation. *Trends Cancer* 2, 505–518.
- Ford, C.E., Jary, E., Ma, S.S.Q., Nixdorf, S., Heinzelmann-Schwarz, V.A., and Ward, R.L. (2013). The Wnt gatekeeper SFRP4 modulates EMT, cell migration and downstream Wnt signalling in serous ovarian cancer cells. *PLoS One* 8, e54362.
- Fresno Vara, J.A., Casado, E., de Castro, J., Cejas, P., Belda-Iniesta, C., and González-Barón, M. (2004). PI3K/Akt signalling pathway and cancer. *Cancer Treat. Rev.* 30, 193–204.
- Ganzfried, B.F., Riester, M., Haibe-Kains, B., Risch, T., Tyekucheva, S., Jazic, I., Wang, X.V., Ahmadifar, M., Birrer, M.J., Parmigiani, G., et al. (2013). curatedOvarianData: clinically annotated data for the ovarian cancer transcriptome. *Database (Oxford)* 2013, bat013.
- Ghosh, A., Syed, S.M., and Tanwar, P.S. (2017). In vivo genetic cell lineage tracing reveals that oviductal secretory cells self-renew and give rise to ciliated cells. *Development* 144, 3031–3041.
- Gilbertson, R.J. (2011). Mapping cancer origins. *Cell* 145, 25–29.
- Ginestet, C. (2011). ggplot2: elegant graphics for data analysis. *J. R. Stat. Soc. Ser. A Stat. Soc.* 174, 245.
- Ginestier, C., Hur, M.H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., Jacquemier, J., Viens, P., Kleer, C.G., Liu, S., et al. (2007). ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell* 1, 555–567.
- Goldman, M., Craft, B., Kamath, A., Brooks, A.N., Zhu, J., and Haussler, D. (2018). The UCSC Xena Platform for cancer genomics data visualization and interpretation. *bioRxiv*, 326470. <https://www.biorxiv.org/content/10.1101/326470v5>.
- Greaves, M., and Maley, C.C. (2012). Clonal evolution in cancer. *Nature* 481, 306–313.
- Gregory, P.A., Bert, A.G., Paterson, E.L., Barry, S.C., Tsykin, A., Farshid, G., Vadas, M.A., Khew-Goodall, Y., and Goodall, G.J. (2008). The mir-200 family and mir-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat. Cell Biol.* 10, 593–601.
- Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., and van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* 525, 251–255.
- Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C., and Lander, E.S. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* 146, 633–644.
- Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646–674.
- Hechenbichler, K., and Schliep, K.P. (2004). Weighted k-Nearest-Neighbor Techniques and Ordinal Classification, Discussion Paper 399, SFB 386 (Ludwig-Maximilians University Munich).
- Hellner, K., Miranda, F., Foto Chedom, D., Herrero-Gonzalez, S., Hayden, D.M., Tearle, R., Artibani, M., KaramiNejadRanbar, M., Williams, R., Gaitskell, K., et al. (2016). Premalignant SOX2 overexpression in the fallopian tubes of ovarian cancer patients: discovery and validation studies. *Ebm* 10, 137–149.
- Honkanen, J., Kainu, T., Ceccatelli, S., Rechardt, L., Hökfelt, T., and Peltola-Huikko, M. (1992). Fos and jun in rat central amygdaloid nucleus and paraventricular nucleus after stress. *Neuroreport* 3, 849–852.
- Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009a). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57.
- Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009b). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13.
- Ince, T.A., Richardson, A.L., Bell, G.W., Saitoh, M., Godar, S., Karnoub, A.E., Iglesias, J.D., and Weinberg, R.A. (2007). Transformation of different human breast epithelial cell types leads to distinct tumor phenotypes. *Cancer Cell* 12, 160–170.
- Ince, T.A., Sousa, A.D., Jones, M.A., Harrell, J.C., Agoston, E.S., Krohn, M., Selfors, L.M., Liu, W., Chen, K., Yong, M., et al. (2015). Characterization of twenty-five ovarian tumour cell lines that phenocopy primary tumours. *Nat. Commun.* 6, 7419.
- Ishay-Ronen, D., Diepenbruck, M., Kalathur, R.K.R., Sugiyama, N., Tiede, S., Ivanek, R., Bantug, G., Morini, M.F., Wang, J., Hess, C., et al. (2019). Gain fat-lose metastasis: converting invasive breast cancer cells into adipocytes inhibits cancer metastasis. *Cancer Cell* 35, 17–32.e6.
- Jiang, Y.-P., Wu, X.-H., Shi, B., Wu, W.-X., and Yin, G.-R. (2006). Expression of chemokine CXCL12 and its receptor CXCR4 in human epithelial ovarian cancer: an independent prognostic factor for tumor progression. *Gynecol. Oncol.* 103, 226–233.
- Kalluri, R. (2016). The biology and function of fibroblasts in cancer. *Nat. Rev. Cancer* 16, 582–598.
- Kang, Y., and Massagué, J. (2004). Epithelial-mesenchymal transitions: twist in development and metastasis. *Cell* 118, 277–279.
- Karatzoglou, A., Smola, A., Hornik, K., and Zeileis, A. (2004). kernlab- an S4Package for kernel methods in R. *J. Stat. Softw.* 11, 1–20.
- Karlan, B.Y., Dering, J., Walsh, C., Orsulic, S., Lester, J., Anderson, L.A., Ginther, C.L., Fejzo, M., and Slamon, D. (2014). POSTN/TGFBI-associated stromal signature predicts poor prognosis in serous epithelial ovarian cancer. *Gynecol. Oncol.* 132, 334–342.
- Kessler, M., Hoffmann, K., Brinkmann, V., Thieck, O., Jackisch, S., Toelle, B., Berger, H., Mollenkopf, H.-J., Mangler, M., Sehouli, J., et al. (2015). The Notch and Wnt pathways regulate stemness and differentiation in human fallopian tube organoids. *Nat. Commun.* 6, 8989.
- Kim, J., Coffey, D.M., Creighton, C.J., Yu, Z., Hawkins, S.M., and Matzuk, M.M. (2012). High-grade serous ovarian cancer arises from fallopian tube in a mouse model. *Proc. Natl. Acad. Sci. U S A* 109, 3921–3926.
- Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161, 1187–1201.
- Klos Dehring, D.A., Vladar, E.K., Werner, M.E., Mitchell, J.W., Hwang, P., and Mitchell, B.J. (2013). Deuterosome-mediated centriole biogenesis. *Dev. Cell* 27, 103–112.
- Koboldt, D.C., Larson, D.E., and Wilson, R.K. (2013). Using VarScan 2 for germline variant calling and somatic mutation detection. *Curr. Protoc. Bioinformatics* 44, 15.4.1–15.4.17.
- Konecny, G.E., Wang, C., Hamidi, H., Winterhoff, B., Kalli, K.R., Dering, J., Ginther, C., Chen, H.-W., Dowdy, S., Cliby, W., et al. (2014). Prognostic and therapeutic relevance of molecular subtypes in high-grade serous ovarian cancer. *J. Natl. Cancer Inst.* 106, 11.
- Koshiyama, M., Matsumura, N., and Konishi, I. (2017). Subtypes of ovarian cancer and ovarian cancer screening. *Diagnostics (Basel)* 7, 12.
- Kuhn, E., Kurman, R.J., Sehdev, A.S., and Shih, I.-M. (2012). Ki-67 labeling index as an adjunct in the diagnosis of serous tubal intraepithelial carcinoma. *Int. J. Gynecol. Pathol.* 31, 416–422.
- Labidi-Galy, S.I., Papp, E., Hallberg, D., Niknafs, N., Adleff, V., Noe, M., Bhattacharya, R., Novak, M., Jones, S., Phallen, J., et al. (2017). High grade serous ovarian carcinomas originate in the fallopian tube. *Nat. Commun.* 8, 1093.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.

- Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 15, R29.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Li, Y.-R., and Yang, W.-X. (2016). Myosins as fundamental components during tumorigenesis: diverse and indispensable. *Oncotarget* 7, 46785–46812.
- Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930.
- Liu, M., Roth, A., Yu, M., Morris, R., Bersani, F., Rivera, M.N., Lu, J., Shioda, T., Vasudevan, S., Ramaswamy, S., et al. (2013). The IGF2 intronic miR-483 selectively enhances transcription from IGF2 fetal promoters and enhances tumorigenesis. *Genes Dev.* 27, 2543–2548.
- Macintyre, G., Goranova, T.E., De Silva, D., Ennis, D., Piskorz, A.M., Eldridge, M., Sie, D., Lewsley, L.-A., Hanif, A., Wilson, C., et al. (2018). Copy number signatures and mutational processes in ovarian carcinoma. *Nat. Genet.* 45, 1127–1129.
- Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161, 1202–1214.
- Mahalingaiah, P.K.S., PonnuSamy, L., and Singh, K.P. (2015). Chronic oxidative stress leads to malignant transformation along with acquisition of stem cell characteristics, and epithelial to mesenchymal transition in human renal epithelial cells. *J. Cell. Physiol.* 230, 1916–1928.
- Marusyk, A., Almendro, V., and Polyak, K. (2012). Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer* 12, 323–334.
- Mateescu, B., Batista, L., Cardon, M., Gruoso, T., de Feraudy, Y., Mariani, O., Nicolas, A., Meyniel, J.-P., Cottu, P., Sastre-Garau, X., et al. (2011). miR-141 and miR-200a act on ovarian tumorigenesis by controlling oxidative stress response. *Nat. Med.* 17, 1627–1635.
- McCarthy, D.J., McCarthy, D.J., Campbell, K.R., Campbell, K.R., Lun, A.T.L., Wills, Q.F., and Wills, Q.F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* 33, 1179–1186.
- McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* 8, 329–337.e4.
- McGranahan, N., and Swanton, C. (2017). Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell* 168, 613–628.
- McLean, G.W., Carragher, N.O., Avizienyte, E., Evans, J., Brunton, V.G., and Frame, M.C. (2005). The role of focal-adhesion kinase in cancer—a new therapeutic opportunity. *Nat. Rev. Cancer* 5, 505–515.
- Meacham, C.E., and Morrison, S.J. (2013). Tumour heterogeneity and cancer cell plasticity. *Nature* 501, 328–337.
- Mora-Castilla, S., To, C., Vaezeslami, S., Morey, R., Srinivasan, S., Dumdie, J.N., Cook-Andersen, H., Jenkins, J., and Laurent, L.C. (2016). Miniaturization technologies for efficient single-cell library preparation for next-generation sequencing. *J. Lab. Autom.* 21, 557–567.
- Nalepa, G., and Clapp, D.W. (2018). Fanconi anaemia and cancer: an intricate relationship. *Nat. Rev. Cancer* 18, 168–185.
- Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453–457.
- Ng, A., Jordan, M., and Weiss, Y. (2002). On Spectral clustering: analysis and an algorithm. NIPS'01 Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic, pp. 849–856.
- Nieto, M.A., Huang, R.Y.-J., Jackson, R.A., and Thiery, J.P. (2016). EMT: 2016. *Cell* 166, 21–45.
- Nowell, P.C. (1976). The clonal evolution of tumor cell populations. *Science* 194, 23–28.
- Obenchain, V., Lawrence, M., Carey, V., Gogarten, S., Shannon, P., and Morgan, M. (2014). VariantAnnotation: a bioconductor package for exploration and annotation of genetic variants. *Bioinformatics* 30, 2076–2078.
- Paik, D.Y., Janzen, D.M., Schafnacker, A.M., Velasco, V.S., Shung, M.S., Cheng, D., Huang, J., Witte, O.N., and Memarzadeh, S. (2012). Stem-like epithelial cells are concentrated in the distal end of the fallopian tube: a site for injury and serous cancer initiation. *Stem Cells* 30, 2487–2497.
- Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344, 1396–1401.
- Perets, R., Wyant, G.A., Muto, K.W., Bijron, J.G., Poole, B.B., Chin, K.T., Chen, J.Y.H., Ohman, A.W., Stepule, C.D., Kwak, S., et al. (2013). Transformation of the fallopian tube secretory epithelium leads to high-grade serous ovarian cancer in Brca1;Tp53;Pten models. *Cancer Cell* 24, 751–765.
- Peters, W.M. (1986). Nature of "basal" and "reserve" cells in oviductal and cervical epithelium in man. *J. Clin. Pathol.* 39, 306–312.
- Phipson, B., Lee, S., Majewski, I.J., Alexander, W.S., and Smyth, G.K. (2016). Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann. Appl. Stat.* 10, 946–963.
- Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 9, 171–181.
- Pils, D., Hager, G., Tong, D., Aust, S., Heinze, G., Kohl, M., Schuster, E., Wolf, A., Sehouli, J., Braicu, I., et al. (2012). Validating the impact of a molecular subtype in ovarian cancer on outcomes: a study of the OVCAD Consortium. *Cancer Sci.* 103, 1334–1341.
- Pisco, A.O., Brock, A., Zhou, J., Moor, A., Mojtabedi, M., Jackson, D., and Huang, S. (2013). Non-Darwinian dynamics in therapy-induced cancer drug resistance. *Nat. Commun.* 4, 2467.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S.R.F., WGS500 Consortium, Wilkie, A.O.M., McVean, G., and Lunter, G. (2014). Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* 46, 912–918.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47.
- Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33, 495–502.
- Schmidt, J.M., Panzilius, E., Bartsch, H.S., Irmler, M., Beckers, J., Kari, V., Linnemann, J.R., Dragoi, D., Hirschi, B., Kloos, U.J., et al. (2015). Stem-cell-like properties and epithelial plasticity arise as stable traits after transient Twist1 activation. *Cell Rep.* 10, 131–139.
- Schwede, M., Waldron, L., Mok, S.C., Wei, W., Basunia, A., Merritt, M.A., Parmigiani, G., Harrington, D., Quackenbush, J., Birrer, M.J., et al. (2018). The impact of stroma on the discovery of molecular subtypes and prognostic gene signatures in serous ovarian cancer. *bioRxiv* 14, 496406. <https://www.biorxiv.org/content/10.1101/496406v1>.
- Seirup, M., Chu, L.-F., Sengupta, S., Leng, N., Shafer, C.M., Duffin, B., Elwell, A.L., Bolin, J.M., Swanson, S., Stewart, R., et al. (2019). Tradeoff between more cells and higher read depth for single-cell RNA-seq spatial ordering analysis of the liver lobule. *bioRxiv*, 764191 <https://www.biorxiv.org/content/10.1101/764191v1>.
- Singh, A., and Settleman, J. (2010). EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. *Oncogene* 29, 4741–4751.
- Soneson, C., and Robinson, M.D. (2018). Bias, robustness and scalability in single-cell differential expression analysis. *Nat. Methods* 18, 255–261.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. *Cell* 177, 1888–1902.e21.

- Svensson, V., da Veiga Beltrame, E., and Pachter, L. (2019). Quantifying the tradeoff between sequencing depth and cell number in single-cell RNA-seq. *bioRxiv*, 762773 <https://www.biorxiv.org/content/10.1101/762773v1>.
- Svensson, V., Vento-Tormo, R., and Teichmann, S.A. (2018). Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* 13, 599–604.
- Tan, T.Z., Miow, Q.H., Huang, R.Y.-J., Wong, M.K., Ye, J., Lau, J.A., Wu, M.C., Bin Abdul Hadi, L.H., Soong, R., Choolani, M., et al. (2013). Functional genomics identifies five distinct molecular subtypes with clinical relevance and pathways for growth control in epithelial ovarian cancer. *EMBO Mol. Med.* 5, 1051–1066.
- Tan, Y., Sinniah, R., Bay, B.H., and Singh, G. (1999). Metallothionein expression and nuclear size in benign, borderline, and malignant serous ovarian tumours. *J. Pathol.* 189, 60–65.
- Thomas, P.D. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 13, 2129–2141.
- Topham, D.J., and Reilly, E.C. (2018). Tissue-resident memory CD8+ T cells: from phenotype to function. *Front Immunol.* 9, 515.
- Torre, L.A., Trabert, B., DeSantis, C.E., Miller, K.D., Samimi, G., Runowicz, C.D., Gaudet, M.M., Jemal, A., and Siegel, R.L. (2018). Ovarian cancer statistics, 2018. *CA Cancer J. Clin.* 68, 284–296.
- Tothill, R.W., Tinker, A.V., George, J., Brown, R., Fox, S.B., Lade, S., Johnson, D.S., Trivett, M.K., Etemadmoghadam, D., Locandro, B., et al. (2008). Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin. Cancer Res.* 14, 5198–5208.
- van Deursen, J.M. (2014). The role of senescent cells in ageing. *Nature* 509, 439–446.
- Visvader, J.E. (2011). Cells of origin in cancer. *Nature* 469, 314–322.
- Wang, S., Chen, J.-Z., Zhang, Z., Huang, Q., Gu, S., Ying, K., Xie, Y., and Mao, Y. (2002). Cloning, characterization, and expression of calcyphosine 2, a novel human gene encoding an EF-hand Ca(2+)-binding protein. *Biochem. Biophys. Res. Commun.* 291, 414–420.
- Wang, Z.A., Mitrofanova, A., Bergren, S.K., Abate-Shen, C., Cardiff, R.D., Califano, A., and Shen, M.M. (2013). Lineage analysis of basal epithelial cells reveals their unexpected plasticity and supports a cell-of-origin model for prostate cancer heterogeneity. *Nat. Cell Biol.* 15, 274–283.
- Wu, R., Zhai, Y., Kuick, R., Karnezis, A.N., Garcia, P., Naseem, A., Hu, T.C., Fearon, E.R., and Cho, K.R. (2016). Impact of oviductal versus ovarian epithelial cell of origin on ovarian endometrioid carcinoma phenotype in the mouse. *J. Pathol.* 240, 341–351.
- Yang, J., Mani, S.A., Donaher, J.L., Ramaswamy, S., Itzykson, R.A., Come, C., Savagner, P., Gitelman, I., Richardson, A., and Weinberg, R.A. (2004). Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* 117, 927–939.
- Yin, X., Fang, S., Wang, M., Wang, Q., Fang, R., and Chen, J. (2016). EFEMP1 promotes ovarian cancer cell growth, invasion and metastasis via activated the AKT pathway. *Oncotarget* 7, 47938–47953.
- Yoshihara, K., Tajima, A., Yahata, T., Kodama, S., Fujiwara, H., Suzuki, M., Onishi, Y., Hatae, M., Sueyoshi, K., Fujiwara, H., et al. (2010). Gene expression profile for predicting survival in advanced-stage serous ovarian cancer across two independent datasets. *PLoS One* 5, e9615.
- Yoshihara, K., Tsunoda, T., Shigemizu, D., Fujiwara, H., Hatae, M., Fujiwara, H., Masuzaki, H., Katauchi, H., Kawakami, Y., Okamoto, A., et al. (2012). High-risk ovarian cancer based on 126-gene expression signature is uniquely characterized by downregulation of antigen presentation pathway. *Clin. Cancer Res.* 18, 1374–1385.
- Zafar, H., Wang, Y., Nakhleh, L., Navin, N., and Chen, K. (2016). Monovar: single-nucleotide variant detection in single cells. *Nat. Methods* 13, 505–507.
- Zhang, Q., Wang, C., and Cliby, W.A. (2019a). Cancer-associated stroma significantly contributes to the mesenchymal subtype signature of serous ovarian cancer. *Gynecol. Oncol.* 152, 368–374.
- Zhang, S., Dolgalev, I., Zhang, T., Ran, H., Levine, D.A., and Neel, B.G. (2019b). Both fallopian tube and ovarian surface epithelium are cells-of-origin for high-grade serous ovarian carcinoma. *Nat. Commun.* 10, 5367.
- Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., and Enard, W. (2017). Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* 65, 631–643.e4.

STAR★METHODS**KEY RESOURCES TABLE**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Cytokeratin 7, clone LP5K, FITC conjugated antibody	Millipore	Cat#CBL194F; RRID: AB_11213981
Anti-CCDC17 antibody produced in rabbit	Sigma-Aldrich	Cat#HPA028338; RRID: AB_10669641
Anti-CAPS antibody produced in rabbit	Sigma-Aldrich	Cat#HPA043520; RRID: AB_10964138
E-Cadherin antibody	BD Biosciences	Cat#610181; RRID: AB_397580
EpCAM (D4K8R) XP® Rabbit mAb antibody	Cell Signaling Technology	Cat#36746; RRID: AB_279910
CD44 Monoclonal Antibody (IM7)	Thermo Fisher Scientific	Cat#14-0441-82; RRID: AB_467246
Op18 (A-4) antibody	Santa Cruz Biotechnology	Cat#sc-48362; RRID: AB_628297
Anti-RGS16 polyclonal antibody	Atlas Antibodies	Cat#HPA055824; RRID: AB_2682934
Anti-SPARC antibody produced in rabbit	Sigma-Aldrich	Cat#HPA003020; RRID: AB_1079531
Anti-KRT17 polyclonal antibody	Atlas Antibodies	Cat#HPA045062; RRID: AB_2679204
Purified anti-human CD3 antibody	BioLegend	Cat#317302; RRID: AB_571927
Pan-Cytokertain (AE1/AE3)	eBioscience	Cat#53-9003-80; RRID: AB_1834350
FITC anti-human CD45 antibody	BioLegend	Cat#304006; RRID: AB_314394
APC anti-human CD103 antibody	BioLegend	Cat#350215; RRID: AB_2563906
HLA-DR antibody (L243)	Gifted from Vincenzo Cerundolo's lab (HIU, WIMM)	N/A
p53 Monoclonal Antibody (DO-7)	Thermo Fisher Scientific	Cat#MA5-12557; RRID: AB_10989883
PAX8 Monoclonal Antibody (1F8-3A8)	Thermo Fisher Scientific	Cat#MA1-117; RRID: AB_2536828
FITC Mouse Anti-Human CD69 Clone FN50	BD Biosciences	Cat#555530; RRID: AB_395915
EpCAM-APC, human	Miltenyi Biotec	Cat#130-113-260; RRID: AB_2726061
Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Thermo Fisher Scientific	Cat#A-11001; RRID: AB_2534069
Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Thermo Fisher Scientific	Cat#A-21202; RRID: AB_141607
Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 546	Thermo Fisher Scientific	Cat#A10040; RRID: AB_2534016
Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 594	Thermo Fisher Scientific	Cat#A-21207; RRID: AB_141637
CD45-FITC, human	Miltenyi Biotec	Cat#130-113-679; RRID: AB_2726220
Mouse Anti-Tubulin IV, beta Monoclonal Antibody	Sigma-Aldrich	Cat#T7941; RRID: AB_261775
Biological Samples		
Surgical specimen from patients with ovarian cancer or endometrial cancer	Department of Gynaecological Oncology, Churchill Hospital, Oxford University Hospitals	N/A
Surgical specimen from benign patients	Oxford University Hospitals	N/A
Chemicals, Peptides, and Recombinant Proteins		
Recombinant RNase Inhibitor	Takara Bio	2313A
RNasin Plus RNase Inhibitor	Promega	Cat#N2611
Fisher Healthcare™ Tissue-Plus™ O.C.T. Compound	Thermo Fisher Scientific	Cat#23-730-571
Triton™ X-100	Sigma-Aldrich	Cat#T8787-50ML
Critical Commercial Assays		
RNAqueous™-Micro Total RNA Isolation Kit	Thermo Fisher Scientific	Cat#AM1931
SMARTer® Stranded Total RNA-Seq Kit v2 - Pico Input	Takara Bio	Cat#634411
KAPA HiFi HS RM	Roche Diagnostics	Cat#07958927001

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
AMPure XP	Beckman Coulter	Cat#A63880
SuperScript™ II Reverse Transcriptase	Thermo Fisher Scientific	Cat#18064014
Tumor Dissociation Kit, human	Miltenyi Biotec	Car#130-095-929
QuantiNova SYBR Green PCR Kit	Qiagen	Cat#208052
Deposited Data		
AOCS	GEO	GSE9899
EMTAB 386	R package CuratedOvarianData (http://bcb.dfci.harvard.edu/ovariancancer/); Ganzfried et al., 2013	N/A
GSE49997	R package CuratedOvarianData (http://bcb.dfci.harvard.edu/ovariancancer/); Ganzfried et al., 2013	N/A
GSE13876	R package CuratedOvarianData (http://bcb.dfci.harvard.edu/ovariancancer/); Ganzfried et al., 2013	N/A
GSE26712	R package CuratedOvarianData (http://bcb.dfci.harvard.edu/ovariancancer/); Ganzfried et al., 2013	N/A
GSE26193	R package CuratedOvarianData (http://bcb.dfci.harvard.edu/ovariancancer/); Ganzfried et al., 2013	N/A
GSE51088	R package CuratedOvarianData (http://bcb.dfci.harvard.edu/ovariancancer/); Ganzfried et al., 2013	N/A
GSE32062.GPL6480	R package CuratedOvarianData (http://bcb.dfci.harvard.edu/ovariancancer/); Ganzfried et al., 2013	N/A
TCGA OvCa	UCSC Xena Data Hub (https://tcga.xenahubs.net); Broad firehose (https://gdac.broadinstitute.org)	Goldman et al., 2018
Single-cell RNA sequencing data of fallopian tubes from cancer patients	GEO	GSE132149
Single-cell RNA sequencing data of fallopian tubes from benign patients	GEO	GSE139079
OXO-PCR bulk RNA sequencing data	GEO	GSE132107
Experimental Models: Cell Lines		
Kuramochi	JCRB	Cat#JCRB0098; RRID: CVCL_1345
Oligonucleotides		
GAPDH-forward (5'-CCCCCACCACTGAATCTC-3')	IDT	N/A
GAPDH-reverse (5'-GGTACATGACAAGGTGCCGC-3')	IDT	N/A
ACTB-forward (5'-GAGCACAGAGCCTCGCCTT-3')	IDT	N/A
ACTB-reverse (5'-ACATGCCGGAGCCGTTGTC-3')	IDT	N/A
TSO (5'-AAGCAGTGGTATCACCGCAGAGTACATGrG+G-3')	Qiagen	N/A
5' biotinylated oligo-dT ₃₀ VN (5' biotin-AAGCAGTGGTA TCAACGCAGAGTACT30VN-3')	IDT	N/A
5' biotinylated ISPCR primer (5' biotin-AAGCAGTGGTA TCAACGCAGAGT-3')	IDT	N/A
Software and Algorithms		
Trim Galore! v0.4.1	http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/	RRID: SCR_011847
STAR v2.4.2a	https://github.com/alexdobin/STAR Dobin et al., 2012	RRID: SCR_015899

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Subread (FeatureCount) v1.4.5	http://subread.sourceforge.net/ Liao et al., 2014	RRID: SCR_009803
R v3.5	https://www.r-project.org	N/A
R package scater v1.10.1	McCarthy et al., 2017	RRID: SCR_015954
R package limma v3.38.3	Law et al., 2014; Phipson et al., 2016; Ritchie et al., 2015	RRID: SCR_010943
R package DoubletFinder v2.0.1	https://github.com/chris-mcginnis-ucsf/DoubletFinder McGinnis et al., 2019	N/A
R package Seurat v3.0.2	Stuart et al., 2019	RRID: SCR_016341
GOrilla: Gene Ontology Enrichment Analysis and Visualization Tool	http://cbl-gorilla.cs.technion.ac.il/ Eden et al., 2007; 2009	RRID: SCR_006848
PANTHER	http://www.pantherdb.org/ Thomas, 2003	RRID: SCR_004869
DAVID v6.8	http://david.abcc.ncifcrf.gov/ Huang et al., 2009a; 2009b	RRID: SCR_001881
R package PhenoPath v1.4.0	Campbell and Yau, 2018	N/A
R package CIBERSORT	https://cibersort.stanford.edu/ Newman et al., 2015	RRID: SCR_016955
R package BSEQ-sc v1.0	Baron et al., 2016	N/A
R package Survival v2.43-3	https://github.com/therneau/survival	N/A
In-house analysis	https://github.com/zhiyhu/scFT-paper	N/A

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Ahmed Ahmed (ahmed.ahmed@wrh.ox.ac.uk). This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS**Cell Line**

KURAMOCHI cell line (cell number ID: JCRB0098) was obtained from Japanese Collection of Research Bioresources (JCRB) Cell Bank. KURAMOCHI cells were cultured in RPMI 1640 media (Gibco) with Fetal Bovine Serum (FBS, 10%; Gibco) and Penicillin-Streptomycin (100 U; Gibco) at 37°C, 5% CO₂ and 95% humidity.

Human Fallopian Tube and Tumor Samples

The cases in this study were recruited under the Gynecological Oncology Targeted Therapy Study 01 (GO-Target-01, research ethics approval #11-SC-0014), the Oxford Radcliffe Biobank (ORB, research ethics approval 09/H0606/5+5 and 19/SC/0173) and the Oxford Ovarian Cancer-Predict Chemotherapy Response Trial (OXO-PCR-01, research ethics approval #12-SC-0404). All participants involved in this study were appropriately informed and consented.

Fallopian tube (FT) biopsies for single-cell RNA-seq were collected from the distal end of fallopian tubes. In total, we collected FT samples from six patients with HGSOC, five patients with endometrial cancer and five patients without cancer conditions ([Table S1](#)). Tumor samples were biopsied during diagnostic laparoscopy, immediately frozen on dry ice and stored in clearly labeled cryovials in -80°C freezers.

METHOD DETAILS**Sample Dissociation**

The fallopian tube samples were processed in 1 hr, after being collected from the hospital. The tissues were dissociated with Human Tumor Dissociation kit (Miltenyi Biotec) and filtered using 100 µm SmarterStrainers (Miltenyi Biotec).

Culture, Cryopreservation and Harvest

Certain samples were cultured or cryopreserved in this study ([Table S1](#)). To culture primary cells from patients, we adopted the medium for FTE culturing ([Kessler et al., 2015](#)) (41.7 ml Advanced DMEM/F12 Medium, 540 µl HEPES, 450 µl P/S, 4.5 µl (10 ng/ml) EGF, 2.25 ml FBS (5%), 45 µl (9 µM) Rock inhibitor/Y0503-5MG, Y-27632 dihydrochloride Sigma-Aldrich, termed as FT culturing medium).

Immediately after dissociation, cells were transferred into a 6-well plate for primary cell culture with the FT culturing medium at 37°C in a CO₂ cell incubator, where they were cultured for overnight or long-term (LT, 2 days or 6 days).

For cryopreservation, cells were suspended in 1 ml Synth-a-Freeze Cryopreservation Medium (Thermo Fisher). Cell suspension was frozen in freezing containers (Thermo Fisher) with Isopropyl alcohol at -80°C for 24 hr before being transferred into liquid nitrogen. The cryopreservation cells were frozen down immediately after dissociation, stored in liquid nitrogen and recovered in FT culturing medium overnight after thawing. We were not able to generate cDNA from single cells that were immediately thawed from liquid nitrogen. Therefore, the frozen cells were recovered overnight in the FT culturing medium before cell sorting.

For cultured cells, cells were harvested with TrypLE (Thermo Fisher) for 5 min at 37°C and dispensed in FACS buffer (1× PBS, 1% BSA, 2 mM EDTA, 0.0025% RNasin plus), and then they were processed as for the freshly dissociated cells.

Organoid Culture

The three-dimensional organoid cultures were conducted as previously described (Kessler et al., 2015). The FT tissue was cut into small pieces of around 2-4 mm in length and incubated in pre-warmed Digestive Medium containing Advanced DMEM/F12 (Gibco), 12 mM HEPES (Gibco), 1% Penicillin/Streptomycin (Gibco), 2 mg/ml Trypsin (Sigma), 0.5 mg/ml DNase I (Sigma) and 100 U/ml Collagenase Type I (Invitrogen); this was incubated for 45 min in rotation at 37°C. Using forceps, decellularized tissue was removed and immediately filtered through a 100 µm cell strainer and pelleted by centrifugation at 300×g for 10 min. The cell pellet was washed once with ice-cold DPBS and resuspended in Matrigel Matrix (Corning). Using an 8-well chamber slide (Thistle Scientific), 20 µl of Matrigel were placed into each well by drops and the mixture was incubated at 37°C for 20 min to allow Matrigel polymerizing and solidifying. Finally, cells were overlaid with pre-warmed organoid medium containing Advanced DMEM/F12, 12 mM HEPES, 1% Penecillin/Streptomycin, 1% N2 Supplement (Thermo Fisher), 2% B-27 Supplement (Thermo Fisher), 100 ng/ml human Noggin (Peprotech), 100 ng/ml human FGF10 (Peprotech), 1 mM Nicotinamide (Sigma), 0.5 µM TGF-β R Kinase Inhibitor VI (SB431542; ThermoFisher), R-Spondin1 protein (RSPO1) and 25% Wnt3a conditioned medium (ATCC). RSPO1 protein was produced in-house using a HA-RSPO1-Fc 293T cell line (Cultrex). Protein expression was performed according to Culturex protocol while protein purification was done using Pierce Protein A Agarose kit (ThermoFisher Scientific).

Cell Sorting

To sort freshly dissociated cells or harvested cells, the cells were dispensed in the FACS buffer (1× PBS, 1% BSA, 2 mM EDTA, 0.0025% RNasin plus) and stained with EpCAM-APC antibody (Miltenyi Biotec) and CD45-FITC antibody (BioLegend) for 30 min at 4°C in the dark. Afterwards, cells were washed once with FACS buffer and kept on ice in the dark. 10 µg/µl DAPI was added 10 min before cell sorting with SH800 sorter (Sony). The alive EpCAM⁺CD45⁻DAPI⁻ fallopian tube epithelial cells were sorted into 96-well plates with one cell per well. Each batch was sorted with bulk controls (10 or 100 cells) and empty controls. A small number of EpCAM⁺CD45⁻DAPI⁻ or EpcAM⁺CD45⁻DAPI⁻ cells were sorted as control for epithelial cells. The cells were sorted into 4 µl lysis buffer with 0.1 µl RNase inhibitor (Clontech), 1.9 µl 0.4% Triton X-100, 1 µl 10 µM 5'-biotinylated oligo-dT₃₀VN (IDT) and 1 µl 10 mM dNTP (Thermo Scientific), snap frozen on dry ice and stored at -80°C before single-cell RNA sequencing.

cDNA Synthesis and Library Construction

The SMART-seq2 protocol (Picelli et al., 2014) was used to generate single-cell cDNA and libraries with optimization and automation. To lyse the cells, the aforementioned lysis buffer with single cells were removed from -80°C, heated at 72°C for 3 min and then kept at 4°C before adding the reverse transcription master mix. The reverse transcription step exactly followed SMART-seq2 protocol with 5'-biotinylated TSO (Qiagen). The first round of 24-cycle PCR started with 10 µl reverse transcription product, 0.125 µl 10 µM 5'-Biotinylated ISPCR primers (IDT), 12.5 µl KAPA HiFi PCR Master Mix (Roche Diagnostics) and 2.375 µl RT-PCR grade water (Invitrogen). PCR products were cleaned up with 0.8:1 Ampure XP beads (Beckman Coulter) with Biomek FxP Laboratory Automation Workstation (Biomek).

The quality of single-cell cDNA was checked on TapeStation with HD5000 Tapes and reagent (Agilent) and single-cell qPCR of GAPDH or ACTB with QuantiNova SYBR Green PCR Kit (Qiagen). The concentration was measured with Quant-iTTM PicoGreen™ dsDNA Assay Kit (Invitrogen) on the CLARIOstar Plate Reader (BMG Labtech). The wells where Ct values of GAPDH or ACTB were less than 20 were defined as the wells with good-quality cDNA. Good-quality cDNA was cherry-picked and normalized to 2 µg/ml with EB (Qiagen) on the Biomek FxP Workstation into a new 384-well plate (4titude).

We used the miniaturized Nextera XT (Illumina) protocol (Mora-Castilla et al., 2016) with Mosquito HTS (tplabtech) to generate libraries from normalized single-cell cDNA in a 384-well Endure plate (Life Technology) with 400 nl input cDNA per reaction. The Nextera index set A and D were used in parallel to multiplex 384 cells per batch. All 384 single-cell libraries in each batch were pooled and cleaned up with 0.6:1 Ampure XP beads (Beckman Coulter). The multiplexed RNA-seq library was normalized to 4 nM with Suspension Buffer (provided in Nextera XT kit) to be sequenced on the NextSeq 500 Sequencer (Illumina).

Whole-Exome Sequencing

The buffy samples and tumor biopsies were stored at -80°C. Tumor biopsies were frozen in the O.C.T. Compound (Fisher Scientific) and sectioned into 1.5 mL tubes for DNA extraction. The tumor sites were confirmed by H&E staining. Genomic DNA (gDNA) was extracted from buffy and tumor samples with the DNeasy Blood & Tissue Kits (Qiagen). The whole-exome sequencing libraries were prepared with the SureSelect Low Input Target Enrichment System (Agilent) and sequenced pair-ended on Illumina HiSeq.

The reads were filtered with Trim Galore!, and mapped with Bowtie2 ([Langmead and Salzberg, 2012](#)). The somatic mutations were called via Platypus ([Rimmer et al., 2014](#)) and VarScan2 ([Koboldt et al., 2013](#)), and then analyzed by R package VariantAnnotation ([Obenchain et al., 2014](#)).

Laser Capture Microdissection

Tissue sectioning, H&E staining and the assessment by a gynecological oncology pathologist were performed as previously described ([Hellner et al., 2016](#)). Tissue biopsies were frozen down with Fisher Healthcare™ Tissue-Plus O.C.T. Compound (NEG-50, Richard-Allan Scientific) and sectioned by 10 µm in CryoStar NX50 Cryostat (Thermo Scientific). The tissue sections were stored at -80°C before proceeding with the staining step.

The pre-chemotherapy tumor biopsies were fresh frozen in OCT and sectioned onto polyethylene naphthalate membrane (PEN) glass slides (MembraneSlide 1.0 PEN, Zeiss) at 6 µm and immediately stored at -80°C until the laser capture microdissection (LCM). Prior to LCM, the slides were taken out from -80°C, immediately immersed in 70% ethanol for 2 min, in Cresyl Violet (Sigma Aldrich) for 2 min, rinsed in 95% ethanol five times to remove residual cresyl violet and dried at room temperature for 2 min.

LCM was conducted on the PALM Laser Microdissection System (Zeiss) and the captured cells were collected onto the adhesive caps of PCR tubes (AdhesiveCap 500 opaque, Zeiss), then immediately resuspended in 10 µl lysis buffer (RNAqueous™-Micro Total RNA Isolation Kit, Invitrogen). RNA was extracted according to the manufacturer's instructions and 3 µl of eluate before DNase treatment was kept for Whole Genome Amplification and stored at -20°C. The remaining RNA was DNase treated, assessed with the Agilent TapeStation 2200 (Agilent) and then stored at -80°C.

Bulk RNA Sequencing

The SMARTer Stranded Total RNA-Seq kit v2 - Pico Input (Takara) was used to prepare sequencing libraries which were indexed, enriched by 15 cycles of amplification, assessed using the Agilent TapeStation and then quantified by Qubit. Multiplexed library pools were quantified with the KAPA Library Quantification Kit (KK4835) and sequenced using 75 bp pair end reads on the Illumina NextSeq500 platform. Fastq files were trimmed for adapter sequences and quality. Trimmed reads were then mapped to the human hg19 reference genome using STAR (v2.4.2a) and read counts were obtained using subread FeatureCounts (v1.4.5-p1).

Immunofluorescent Staining

To do immunofluorescent staining on the frozen OCT-embedded tissue sections, slides were taken out from -80°C and thawed at room temperature for 5 min. The slides were washed with ice-cold PBS twice to remove the OCT and then fixed in ice-cold 4% PFA for 10 min. The fixed slides were washed with PBS/Glycine buffer to remove residual PFA and then permeabilized with Permeabilization Buffer (1× PBS, 0.5% Triton X-100) for 10 min at RT. A hydrophobic circle was drawn around the tissue region with the PAP pen (Abcam) and dried for 5 min at room temperature (RT). The permeabilized slides were washed in Wash Buffer (1× PBS, 0.2% Triton X-100, 0.05% Tween-20) and blocked in Blocking Buffer (Wash buffer, 10% donkey serum, Sigma) for 1-2 hr. The diluted primary antibodies were incubated at 4°C overnight or at RT for 1-2 hr. The slides were washed and incubated with the secondary antibodies (1:200) and DAPI (1:1000) at RT for 1-2 hr. The stained slides were washed, mounted with the mounting medium without DAPI (Vector Laboratories) and covered. Then they were dried for 1-2 hr at RT in dark before being imaged using a confocal microscope (Zeiss 780). The samples used for staining are listed in [Table S1](#).

Immunohistochemistry

The Formalin-Fixed Paraffin-Embedded (FFPE) samples were sectioned into 2.5 or 4 µm. We used the Leica Bond Max autostainer (Leica Microsystems) or Autostainer plus Link 48 (Dako) to automatically stain the tissue sections. The standard Heat Induced Epitope Retrieval (HIER) in retrieval buffer pH 6 was used. Tissue sections were kept at 100°C for 20 min and incubated with the primary antibody or IgG control for up to 1 hr. We used the Aperio slide scanner (Aperio) to scan the stained slides at 20× or 40× magnification. The samples used for staining are listed in [Table S1](#).

QUANTIFICATION AND STATISTICAL ANALYSIS

Preprocessing scRNA-seq Data

Sequencing files were trimmed by Trim Galore!, mapped to UCSC hg19 human genome assembly by STAR ([Dobin et al., 2012](#)) and counted by FeatureCount ([Liao et al., 2014](#)). The quality of the expression data was checked by the R package scater ([McCarthy et al., 2017](#)) and the low-quality cells were filtered out from further analysis. We used three criteria to filter cells for further analysis: first, the number of genes detected per cells was more than 1200 and less than 7500; second, the total number of read counts was more than 0.1 million per cell; third, the percentage of counts of the top highly expressed 200 genes was less than 85% to ensure enough complexity of the transcriptomes. The upper cutoff for numbers of genes per cells was aimed to filter out the putative doublets. After filtering, the numbers of detected genes per cell were between 1,202 and 7,496 (mean: 3,665). The numbers of total reads per cell were between 102,319 and 9,668,129 (mean: 763,233).

Detecting Doublets

To effectively detect doublets, we used R package DoubletFinder ([McGinnis et al., 2019](#)). In brief, we combined all fresh cells (secretory cells, ciliated cells, fibroblasts and leukocytes), constructed a Seurat S4 object ([Butler et al., 2018](#)) and ran DoubletFinder with its default pipeline.

Visualization of Expression Data

The uniform manifold approximation and projection (UMAP) plots were based on the functions of runUMAP and plotUMAP, and the violin plots of expression were plotted by the function plotExpression from the R package scater. Heatmaps were plotted by R packages pheatmap and ggplot2 ([Ginestet, 2011](#)).

Comparing Fresh and Cultured Cells

To investigate the effects of cell culture, we first combined secretory cells from Patient 11553 and Patient 15072. Then we used limma-voom to identify the DE genes in three pairs of comparisons, namely fresh versus overnight, fresh versus LT and overnight versus LT. The threshold of DE genes was fold-change ≥ 2 and FDR < 0.05 . The total number of DE genes affected by culture condition was 4061. The GO analysis was conducted by GOrilla with cutoffs at FDR < 0.05 .

Pseudotime Analysis

We used the R package PhenoPath ([Campbell and Yau, 2018](#)) to perform the pseudotime analysis that projected the high-dimensional transcriptomic data to one dimension, in which we compared fresh, overnight cultured and long-term cultured cells. The expression matrix that was normalized by the R package scater was used as the input and the first principal component was used as the initialization of the latent trajectory to accelerate the convergence.

Filtering to Select Fresh Secretory Cells

Only the cells that belonged to the secretory cluster and the fresh group were kept except for the cells from the 15066-left tube (15066L), because the cells from 15066L formed a patient-specific cluster ([Figure 1F](#)) and it was reported to have salpingitis on this side. Secretory cells with strong expression of *KRT7* and *EPCAM* ($\log [CPM+1] > 2$), no expression for *PTPRC* ($\log [CPM+1] = 0$) or *CCDC17* ($\log [CPM+1] < 1$) and no observed copy number variants were kept for the clustering analysis. Copy number variants were detected by HoneyBADGER with default settings ([Fan et al., 2018](#)).

Single-nucleotide variants (SNVs) were called from piled up BAM files by MonoVar ([Li et al., 2009; Zafar et al., 2016](#)) with the recommended parameters (samtools mpileup -BQ0 -d10000 -f ref.fa -q 40 -b filenames.txt | monovar.py -p 0.002 -a 0.2 -t 0.05 -m 2 -f ref.fa -b filenames.txt -o output.vcf). SNVs in exome regions were intersected by the Bedtools ([Quinlan and Hall, 2010](#)).

Differentially-Expressed-Based Clustering

We used an in-house clustering algorithm to identify the cell populations across patients. In brief, the clustering method contained three steps: (1) an initial clustering step within the individual patients by Spectral Clustering ([Karatzoglou et al., 2004; Ng et al., 2002](#)); (2) calculation of distance between each pair of initial clusters by differential expression (DE) analysis; (3) a final hierarchical clustering across different patients based on the calculated distance matrix ([Figure 1B](#)).

The clustering approach for clinical scRNA-seq data required the input of an expression matrix, where the rows corresponded to genes/features and the columns corresponded to cells. The patient source of all cells was denoted by a vector. The length of this source vector was equal to number of cells and the number of different values in the source vector was equal to the number of patients. For the dataset containing multiple patients, our method consisted of three main steps.

The count matrix was preprocessed by log-normalization, selection of high variance genes ([Satija et al., 2015](#)), centered and dimensionally reduced by the principal component analysis. The first step was initial clustering of the cells within individual patients by the weighted k-nearest neighbors spectral clustering provided by knn ([Hechenbichler and Schliep, 2004](#)). After spectral clustering, each cell was assigned to a patient-specific initial cluster. Cells from different patients were clustered into separate patient-specific clusters.

The second step was pairwise differential expression (DE) analysis by limma voom ([Law et al., 2014; Phipson et al., 2016; Ritchie et al., 2015](#)) between initial clusters across patients to build a distance matrix, because it was reported to be one of the top ranking DE analysis methods for scRNA-seq ([Soneson and Robinson, 2018](#)). There were two modes to calculate the distance. The first one was unweighted, meaning that the distance between two initial clusters was defined as the number of differentially expressed genes (DEGs) between them. The second mode was weighted. The DE genes had different weights in the distance. The DE genes with higher fold change and larger ratio of expression had higher weight. The weighted method had advantages in identifying rare but functionally distinguished cell types. As a result, a distance matrix of functional dissimilarity was constructed among all initial clusters.

In the last step, all initial clusters were assigned to the final clusters by hierarchical clustering based on the aforementioned distance matrix. Accordingly, the single cells were assigned to the final cluster containing the initial cluster that the cells belong to. The hierarchical tree was plotted to help dividing the final clusters.

Data Integration

To validate the discovered secretory subtypes in the samples from benign patients, we used the data integration from Seurat v3 ([Stuart et al., 2019](#)). The integration step merged the high-quality fresh secretory cells from five cancer patients and five benign patients.

The cells were first split by their batches. We normalized the count matrix by the vst method and selected the top 1000 high variance genes by using Seurat. Then we computed the integration anchors by using the first 10 dimensions and set the number of k nearest neighbors for finding anchors as 5 (other parameters were default), and then we merged the data by using the computed anchors.

Identifying Marker Genes

The identification of markers was based on limma-voom, which has good performance with scRNA-seq data (Soneson and Robinson, 2018). To spot the positive markers in the target cluster, we conducted the differential expression analysis between the target cluster and background (non-target) clusters. Only if one gene emerged in all the pairwise comparisons between the target cluster and background clusters with FDR < 0.05 and \log_2 fold-change > 0.4, this gene was defined as the marker gene of this target cluster.

Gene Ontology Enrichment Analysis

Gene ontology enrichment analysis was conducted on GOrilla (last updated on May 26, 2018) (Eden et al., 2007, 2009). The statistical overrepresentation was conducted by the PANTHER Gene list analysis (Thomas, 2003). Pathway enrichment analysis was done by the DAVID Bioinformatics Resources 6.8 (Huang et al., 2009a; 2009b). The threshold of significance was FDR < 0.05.

Bulk Expression Data

TCGA data was downloaded from the UCSC Xena Data Hub (<https://tcga.xenahubs.net>) (Goldman et al., 2018) and the Broad firehose (<https://gdac.broadinstitute.org>), which was derived from the TCGA Data Coordinating Center (version: Jan 2016). We used the “IlluminaHiSeq UNC” RNA-seq dataset (version: 2017-10-13) and the “TCGA OV gene expression subtype” phenotype data (version: 2016-05-27). The AOCS dataset was downloaded from GSE9899. The other microarray datasets were retrieved from the R package CuratedOvarianData (Ganzfried et al., 2013).

Deconvolution of Bulk Expression Data

On the basis of our FTE scRNA-seq data, we firstly selected the marker genes for five FTE subtypes, namely Differentiated (C3), KRT17 cluster (C4), Cell cycle (C9), EMT (C7) and Ciliated signatures by using the following cutoffs: (1) \log_2 fold-change ≥ 1.5 ; (2) FDR ≤ 0.01 ; (3) dispersion ≥ 0.2 that was measured by the variable loadings in the principal component analysis of TCGA RNA-seq data; (4) the correlation with each other markers was less than 0.9 in TCGA data to avoid redundancy. We calculated the reference matrix (Figure 5A and Table S7) by BSEQ-sc (Baron et al., 2016) that measured that average expression levels of marker genes in each cell subtype. We used this reference matrix to deconvolute the RNA-seq data from TCGA, the microarray data from AOCS and CuratedOvarianData and the OXO-PCR RNA-seq data by CIBERSORT as previously described (Newman et al., 2015), which is based on the v-parameter support vector regression and is robust to noise and outliers. The CIBERSORT was ran in the relative mode with non-logged expression matrices. For the microarray data, if a gene corresponded to multiple probes, we selected the probe that had the highest average expression level to represent that gene. The deconvolution analysis generated scores of the five transcriptomic signatures in each tumor sample (Figure 5A). This score of a signature can be interpreted as the proportion of the corresponding cell state in a tumor sample.

Survival Analysis

The survival analysis was conducted by fitting a Cox proportional hazards regression model. This analysis used age, grade and residual disease as covariates. We repeated the survival analysis in TCGA RNA-seq dataset, AOCS microarray dataset and other seven microarray datasets from the CuratedOvarianData database (Ganzfried et al., 2013). The criterion of selecting datasets was that a dataset contained over one hundred clinical samples. We filtered out the non-serous cases. The following analysis was performed on SOC cases only. For the datasets from CuratedOvarianData that contained both low and high grades or early and late stages, we used the multivariate analysis to eliminate the effects of confounding factors. In the multi-variate survival analysis, we dichotomized grades to low-grade (Grade 1) and high-grade (Grades 2 and 3) and dichotomized stages to early (Stages I and II) and late (Stages III and IV) (Table S9). EMT scores and overall survival were negatively correlated in the independent eight datasets ($p < 0.05$), except for the “GSE32062.GPL6480” dataset (see Table 1). The hazard ratios of EMT scores in all the nine independent datasets were larger than 1 (range: 1.89 – 3.11). For the survival curve of the meta-analysis (Figure 6C), we dichotomized the EMT-low and EMT-high groups by the median of EMT scores of each dataset to avoid batch effects.

DATA AND CODE AVAILABILITY

The code generated during this study are available at GitHub (web link: <https://github.com/zhiyhu/scFT-paper>). Single-cell RNA sequencing datasets of fallopian tubes from cancer and benign patients are deposited at Gene Expression Omnibus (GEO accession numbers: GSE132149 and GSE139079, respectively). The bulk RNA sequencing dataset is deposited at Gene Expression Omnibus (GEO accession number: GSE132107).