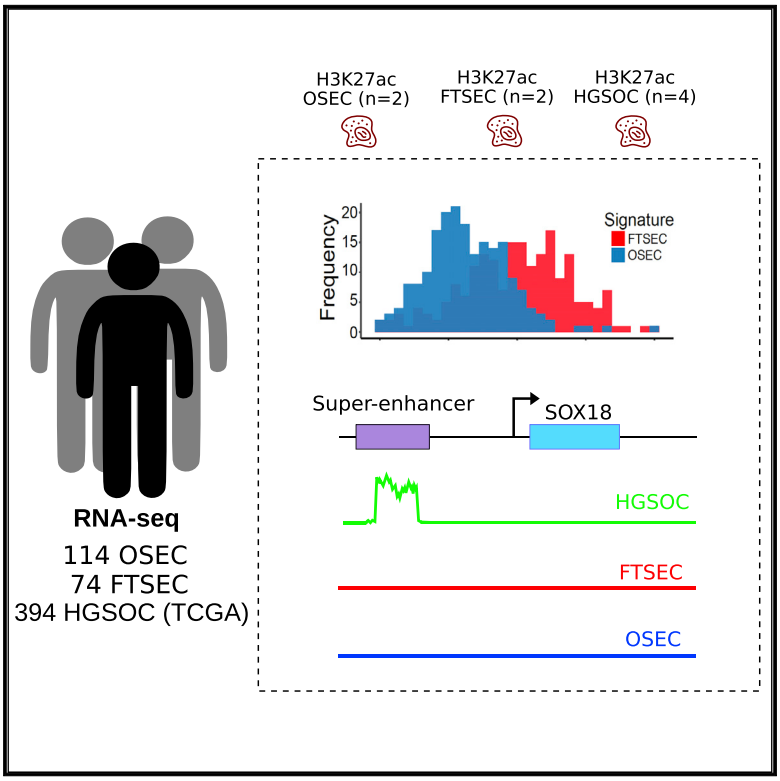
Resource

Cell Reports

A Study of High-Grade Serous Ovarian Cancer Origins Implicates the SOX18 Transcription Factor in Tumor Development

Graphical Abstract



Highlights

- The majority of high-grade serous ovarian cancers arise from fallopian epithelia
- A subset of proliferative-type tumors likely arise from ovarian epithelial cells
- Super enhancers dysregulate transcription factor expression during tumorigenesis
- SOX18 induces an epithelial-to-mesenchymal transition in fallopian tube epithelia

Authors

Kate Lawrenson, Marcos A.S. Fonseca, Annie Y. Liu, ..., Matthew L. Freedman, Simon A. Gayther, Houtan Noushmehr

Correspondence

kate.lawrenson@cshs.org (K.L.), simon.gayther@cshs.org (S.A.G.), hnoushm1@hfhs.org (H.N.)

In Brief

Lawrenson et al. profile gene expression and active chromatin in ~200 ovarian and fallopian epithelial isolates and implement machine learning to demonstrate that most high-grade serous ovarian cancers (HGSOCs) derive from fallopian tube epithelial cells, but a subset may originate from ovarian epithelia. SOX18 induces mesenchymal features to drive early neoplasia in fallopian tube precursors.

Lawrenson et al., 2019, Cell Reports *29*, 3726–3735
December 10, 2019
https://doi.org/10.1016/j.celrep.2019.10.122

CellPress

OPEN ACCESS
CellPress

Cell Reports
Resource

A Study of High-Grade Serous Ovarian Cancer Origins Implicates the SOX18 Transcription Factor in Tumor Development

Kate Lawrenson,[1,2,14,*] Marcos A.S. Fonseca,[1,14] Annie Y. Liu,[1,14] Felipe Segato Dezem,[1,14] Janet M. Lee,[2] Xianzhi Lin,[1] Rosario I. Corona,[1,2] Forough Abbasi,[1,2] Kevin C. Vavra,[1,2] Huy Q. Dinh,[3] Navjot Kaur Gill,[4] Ji-Heui Seo,[5] Simon Coetzee,[2] Yvonne G. Lin,[6,15] Tanja Pejovic,[6,7,8] Paulette Mhawech-Fauceglia,[9] Amy C. Rowat,[4] Ronny Drapkin,[10] Beth Y. Karlan,[1] Dennis J. Hazelett,[2] Matthew L. Freedman,[5] Simon A. Gayther,[2,13,16,*] and Houtan Noushmehr[11,12,13,*]

[1]Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA
[2]Center for Bioinformatics and Functional Genomics, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA
[3]Division of Inflammation Biology, La Jolla Institute for Immunology, La Jolla, CA, USA
[4]Department of Integrative Biology and Physiology, University of California, Los Angeles, Los Angeles, CA, USA
[5]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA
[6]Division of Gynecologic Oncology, Department of Obstetrics-Gynecology, University of Southern California/Keck School of Medicine, Los Angeles, Los Angeles, CA, USA
[7]Department of Obstetrics and Gynecology, Oregon Health and Science University, Portland, OR, USA
[8]Knight Cancer Institute, Oregon Health & Science University, Portland, OR, USA
[9]Aurora Diagnostics, Austin, TX, USA
[10]Penn Ovarian Cancer Research Center, Department of Obstetrics and Gynecology, University of Pennsylvania, Philadelphia, PA, USA
[11]Department of Neurosurgery, Henry Ford Hospital, Detroit, MI, USA
[12]Department of Genetics, Ribeirão Preto Medical School, University of São Paulo, São Paulo, Brazil
[13]Senior author
[14]These authors contributed equally
[15]Present address: Genentech, South San Francisco, CA, USA
[16]Lead Contact
*Correspondence: kate.lawrenson@cshs.org (K.L.), simon.gayther@cshs.org (S.A.G.), hnoushm1@hfhs.org (H.N.)
https://doi.org/10.1016/j.celrep.2019.10.122

SUMMARY

Fallopian tube secretory epithelial cells (FTSECs) are likely the main precursor cell type of high-grade serous ovarian cancers (HGSOCs), but these tumors may also arise from ovarian surface epithelial cells (OSECs). We profiled global landscapes of gene expression and active chromatin to characterize molecular similarities between OSECs (n = 114), FTSECs (n = 74), and HGSOCs (n = 394). A one-class machine learning algorithm predicts that most HGSOCs derive from FTSECs, with particularly high FTSEC scores in mesenchymal-type HGSOCs ($p_{adj} < 8 \times 10^{-4}$). However, a subset of HGSOCs likely derive from OSECs, particularly HGSOCs of the proliferative type ($p_{adj} < 2 \times 10^{-4}$), suggesting a dualistic model for HGSOC origins. Super-enhancer (SE) landscapes were also more similar between FTSECs and HGSOCs than between OSECs and HGSOCs ($p < 2.2 \times 10^{-16}$). The SOX18 transcription factor (TF) coincided with a HGSOC-specific SE, and ectopic overexpression of SOX18 in FTSECs caused epithelial-to-mesenchymal transition, indicating that SOX18 plays a role in establishing the mesenchymal signature of fallopian-derived HGSOCs.

INTRODUCTION

Invasive epithelial ovarian cancers are a heterogenous group of tumors comprising several major histological subtypes: high-grade serous, low-grade serous, endometrioid, clear cell, and mucinous. High-grade serous ovarian cancer (HGSOC) is the most common subtype, comprising around two-thirds of all invasive cases. Our understanding of the cellular origins of HGSOC and key transcription factor networks deregulated during HGSOC development has been restricted by the lack of substantial molecular profiling data for the putative precursor tissues, specifically fallopian tube secretory epithelial cells (FTSECs) and ovarian surface epithelial cells (OSECs).

Historically, HGSOCs were thought to arise from OSECs, an atypical epithelial cell type with mesothelial features and inherent phenotypic plasticity and heterogeneity (Kruk and Auersperg, 1992; Park et al., 2018). However, examples of early-stage ovarian carcinoma arising from OSECs *in vivo* are rare. The discovery of occult carcinomas in the fallopian tubes of *BRCA1* and *BRCA2* mutation carriers supports an alternative hypothesis that the fallopian epithelium harbors the cell-of-origin for HGSOC (Callahan et al., 2007; Leeper et al., 2002; Medeiros et al., 2006; Paley et al., 2001; Piek et al., 2001). Subsequent studies have shown that a substantial proportion of all HGSOC cases in non-*BRCA1/2* mutation carriers arise from the fallopian tube and, more specifically,

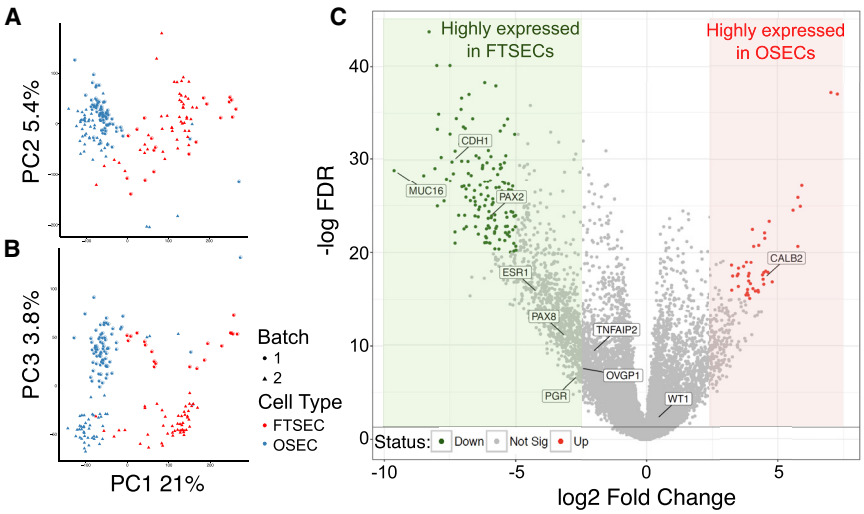


Figure 1. Transcriptomic Profiling of OSECs and FTSECs

(A and B) Principal-component analysis (PCA) of RNA-seq profiles of OSECs (n = 114) and FTSECs (n-74). OSEC samples tend to cluster more tightly together, whereas FTSEC samples show more diffuse clustering. PCA analyses were divided into dimensions 1 and 2 (A) and dimensions 1 and 3 (B). This suggests greater inter-patient heterogeneity between for FTSEC samples.

(C) Volcano plot illustrating differential gene expression between OSEC and FTSEC samples. Known cell-type-specific markers for each cell type are indicated.

the tubal secretory epithelial cells (Gilks et al., 2015; Kindelberger et al., 2007; Labidi-Galy et al., 2017). However, there is no evidence of fallopian tube involvement in other cases, suggesting that other cell types may be precursors for a proportion of HGSOCs.

The goal of this study was to investigate the hypothesis that FTSECs and OSECs both represent cells of origin of HGSOC. To do this, we compared the molecular relationships between OSECs, FTSECs, and HGSOCs based on transcriptomic and epigenomic profiles. We first used machine learning to identify transcriptional signatures of disease origins by using data from 114 OSECs, 74 FTSECs, and 394 HGSOCs. We then performed chromatin immunoprecipitation sequencing (ChIP-seq) to map active chromatin in OSECs, FTSECs, and HGSOCs and characterize tissue-specific super-enhancer landscapes. Finally, we integrated ChIP-seq and transcriptomic data to identify transcription factors SOX18, ELF3, and EHF as putative drivers of transcriptional deregulation in HGSOC development. Characterizing the exact origins of the HGSOC will be essential for the development of effective tumor prevention and early detection strategies in the future.

RESULTS

Expression Profiling of Putative Ovarian Cancer Precursor Cells

One approach to investigate the cellular origins of cancer is to quantify similarities and differences between molecular signatures of tumors and the proposed tissues of origin (Staub et al., 2010). This is based on the hypothesis that the molecular blueprint of normal precursor cells is maintained in developing tumors. We performed RNA sequencing (RNA-seq) in 74 FTSEC and 114 OSEC short-term cultures established from 132 individuals; OSEC and FTSEC specimens were derived from the same individual in 56 cases (Karst et al., 2011; Lawrenson et al., 2009) (Table S1). To ensure quality control, we performed RNA-seq in duplicate for five samples (one OSEC and four FTSEC specimens), which confirmed high correlation of expression profiles across between replicates (Pearson's correlation r = 0.79–0.98) (Figure S1A). We found no associations with experimental or epidemiological variables (where available), including sample preparation, patient age, or patient ethnicity (data not shown).

We used principal-component analysis (PCA) to compare expression profiles of OSECs and FTSECs. The two cell types largely stratified according to their molecular profiles (Figures 1A and 1B). We identified 87 significantly differentially expressed genes (DEGs) between OSECs and FTSECs (absolute $\log_2$ fold change [FC] > 2, $p_{adj} = 10^{-30}$; Figure 1C; Table S2). These included *MUC16* (which encodes ovarian cancer screening marker CA125) and *CDH1* (E-cadherin), two genes already known to be differentially expressed between these cell types (Figures 1C and S1B). We also identified overexpressed genes in OSECs; these included *GATA4* (FC = 7.1, $p_{adj} = 3.78 \times 10^{-42}$) and *NR5A1* (FC = 6.7, $p_{adj} = 2.59 \times 10^{-39}$) both of which are transcriptional activators potentially involved in the differentiation of OSECs. Differentially expressed genes (DEGs) that are highly expressed in FTSECs compared to OSECs include genes that encode the cell surface or secreted proteins *MMP7* (FC = 9.9, $p_{adj} = 1.87 \times 10^{-31}$), *CLIC5* (FC = 8.36, $p_{adj} = 5.14 \times 10^{-49}$), *TACSTD2* (FC = 8.21, $p_{adj} = 1.2 \times 10^{-42}$), and *CFTR* (FC = 8.15, $p_{adj} = 7.35 \times 10^{-31}$).

Machine Learning to Predict Cell of Origin for HGSOCs

We applied machine learning algorithms to predict the cell of origin for 394 primary HGSOCs profiled by The Cancer Genome Atlas (TCGA). To correct for differences in read depth and RNA-seq methods between studies, we aligned, batch corrected, and normalized all three datasets—OSEC, FTSEC, and TCGA—together (see STAR Methods). We first defined cell-type-specific signatures of OSECs and FTSECs and then applied a one-class logistic regression (OCLR) methodology, which is particularly well suited to scenarios where a negative class cannot be clearly defined (Sokolov et al., 2016a). First, we tested the performance of the models in identifying OSECs mixed into an FTSEC background and vice versa. Area under the curve (AUC) statistics generated using a leave-one-out approach indicated that the OCLR models performed with

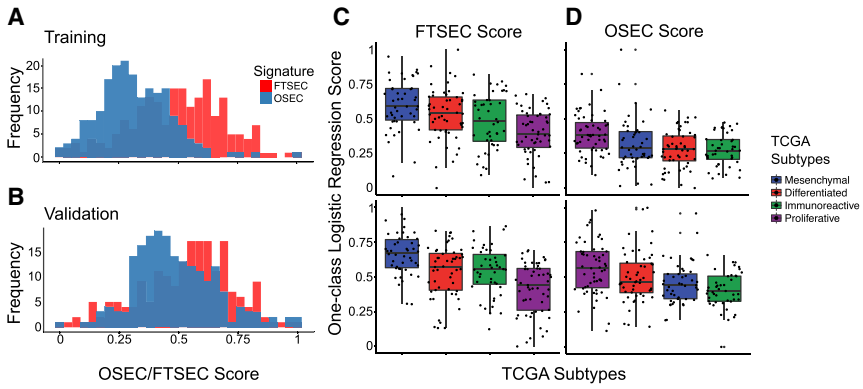


Figure 2. One-Class Logistic Regression (OCLR) Predictors of the Cellular Origins of HGSOC

(A and B) OSEC and FTSEC whole-transcriptomic signatures were developed and compared to whole-transcriptomic signatures of 394 primary HGSOCs publicly available from TCGA project. TCGA analyses were divided into test (n = 197) (A) and validation (n = 197) (B) sets of tumors. In both test and validation sets, FTSEC score tended to be higher in HGSOCs than OSEC scores. The dashed line indicates a score of 0.5.

(C and D) FTSEC (C) and OSEC (D) signatures were compared across HGSOC molecular subgroups. FTSEC score was highest in the mesenchymal subgroup in both test and validation sets; OSECs were highest in the proliferative subgroup.

high specificity (average AUC for OSECs = 0.99 and for FTSECs = 0.97). OCLR models provide a score for each sample and for each category, which is rescaled between zero and one, where zero implies no similarity and one implies high similarity. We applied the OCLR models to HGSOCs to generate an OSEC and FTSEC score for each individual tumor to determine which cell type represents the most likely cell of origin. HGSOC samples were randomized and divided into two equally sized groups (n = 197), designated the training set and the validation set. Each set included similar numbers of the four HGSOC molecular subgroups—differentiated, immunoreactive, mesenchymal, and proliferative—classified by their gene expression signatures (Cancer Genome Atlas Research Network, 2011; Tothill et al., 2008). In both the training and validation datasets, we observed a greater proportion of HGSOCs with higher FTSEC scores than OSEC scores. In the training set, 103/197 tumors (52%) had an FTSEC score > 0.5, whereas only 20/197 tumors (10%) had an OSEC score > 0.5. In the validation set, 124/197 tumors (63%) and 82/197 tumors (42%) had FTSEC and OSEC scores > 0.5, respectively (Figures 2A and 2B). Taken together, these data indicate that across the whole dataset, transcriptome signatures of HGSOCs are more similar to those of FTSECs than OSECs, consistent with a large body of data indicating that FTSECs are the most common cell of origin for HGSOC. There was a weak negative correlation between tumor FTSEC and OSEC scores (Figure S2A) (Pearson's product-moment correlation = −0.16, p = 0.002). In a PCA performed using all expressed genes, FTSECs cluster more closely to HGSOCs than OSECs (Figure S2B). However, 19 tumors (4.8% of cases) had OSEC scores greater than 0.75, indicating they have most likely derived from ovarian surface epithelial cells. Taken together, these data are consistent with the hypothesis that HGSOCs can originate from both FTSECs and OSECs, with FTSECs the most common cell of origin (Eckert et al., 2016; Pothuri et al., 2010; Salazar et al., 1996).

We also investigated if FTSEC and OSEC scores correlate with clinical and molecular features of HGSOCs. In both the training and validation datasets, mesenchymal-type HGSOCs had significantly higher FTSEC OCLR scores ($p_{adj} < 0.02$ in the training and validation cohort; Figure 2C; $p_{adj} = 8 \times 10^{-4}$ in a meta-analysis of all 394 HGSOCs); patients in this subgroup of HGSOC had the worst survival (Cancer Genome Atlas Research Network, 2011). By contrast, patients with proliferative-type HGSOCs had significantly larger OSEC scores ($p_{adj} < 0.001$ in the training and validation cohort, $p_{adj} = 2 \times 10^{-4}$ in a meta-analysis) (Figure 2D), indicating that OSEC-derived tumors are enriched in this molecular subgroup. Finally, we tested for associations between FTSEC and OSEC OCLR scores and patient age, tumor stage, tumor grade, chemoresponse, and debulking status (Figures S2C and S2D). We found no significant associations for FTSECs, but tumors with high OSEC scores were associated with older age at diagnosis ($p_{adj} = 0.005$, normalized enrichment score = 1.6). Higher OSEC score was modestly associated with increased sensitivity to chemotherapy ($p_{adj} = 0.03$, normalized enrichment score = 1.5) (Figure S2D).

FTSEC Super-Enhancer Landscapes Are Conserved in HGSOCs

Epigenomic signatures can also serve as indicators of cell lineage. We used chromatin immunoprecipitation sequencing (ChIP-seq) for H3K27ac to characterize epigenomic landscapes in OSECs (n = 2), FTSECs (n = 2), and HGSOCs (n = 4), and using these data super-enhancer (SE) landscapes, defined as dense clusters of highly active chromatin that typically localize with master regulators of cellular identity (Whyte et al., 2013). OSECs and HGSOCs had the largest numbers of cell-type-specific SEs (n = 337 and n = 336, respectively). Significantly more SEs were shared between FTSECs and HGSOCs (n = 80) than between OSECS and HGSOCs (n = 37) (odds ratio = 12.9, Fisher's exact test, $p < 2.2 \times 10^{-16}$; Figures 3A and 3B). Using the transcriptomic data shown in Figure 1, we verified tissue-specific overexpression of genes proximal to tissue-specific SEs (Figures 3C, 3F, and 3I). The *PAX8* transcription factor (TF) was overexpressed in both FTSECs and HGSOCs and coincides with a SE detected in both cell types at this locus; PAX8 is a well-established biomarker that is ubiquitously expressed in FTSECs and is overexpressed in the majority of primary HGSOCs (Cheung et al., 2011; Laury et al., 2011; Mhawech-Fauceglia et al., 2012) (Figure S3). Supporting this, using PAX8 ChIP-seq data in ovarian cancer cell lines and FTSECs, we observed PAX8 binding within the PAX8 super enhancer (Figure S3). We also identified candidate genes regulated by SEs in each cell type, including *SULT1B1* in OSECs (Figures 3D and 3E) and the

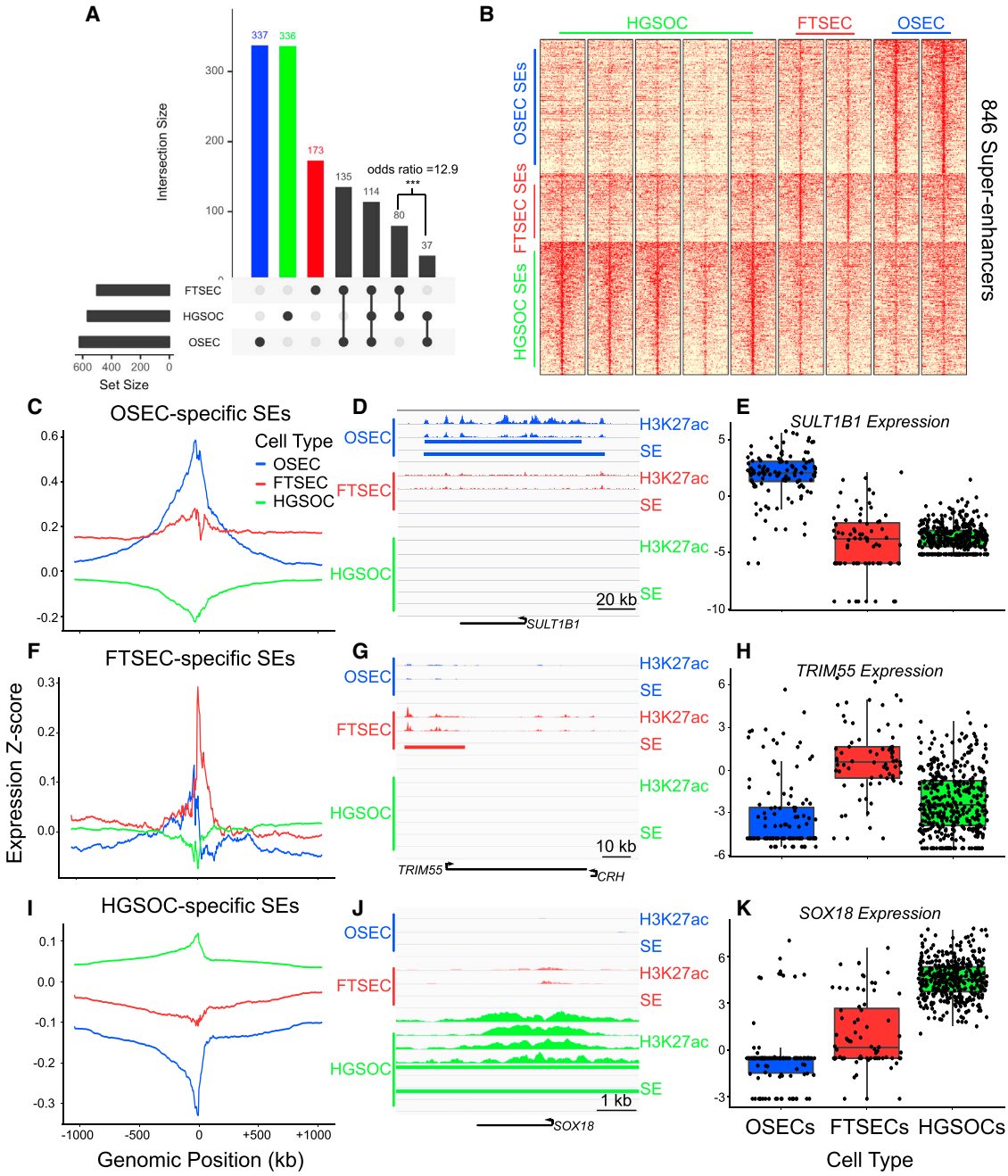


Figure 3. Super-Enhancer-Gene Relationships in OSECs, FTSECs, and HGSOCs

(A) UpSetR (pseudo-venn) diagram of the SE catalog from OSECs (n = 2), FTSECs (n = 2), and HGSOCs (n = 4), showing SE intersections across the three tissue types. Although fewer SEs were cataloged in FTSECs than in OSECs and HGSOCs, significantly more SEs are shared between FTSECs and HGSOCs than between OSECs and HGSOCs (Fisher's exact test, $p < 2.2 \times 10^{-16}$).

(B) The landscape of cell-type-specific SEs across OSECs, FTSECs, and HGSOCs. For OSECs and FTSECs, H3K27ac ChIP-seq data generated for two independent immortalized normal lines per cell type were used to identify SEs. For HGSOCs, H3K27ac ChIP-seq data were generated for four different primary HGSOCs.

(C–K) Tissue-specific SEs associated with elevated gene expression in *cis* in a cell-type-specific manner.

(C–E) OSEC-specific SEs.

(F–H) FTSEC-specific SEs.

(I–K) HGSOC-specific SEs.

(legend continued on next page)

Tripartite Motif Containing 55 (*TRIM55*) gene in FTSECs (Figures 3G and 3H). The *SOX18* transcription factor was marked by an SE in HGSOCs but not in normal tissues, and SOX18 was overexpressed in tumors, suggesting this is a SE-driven TF in HGSOC (Figures 3J and 3K).

SOX18 Is a Driver of Epithelial-to-Mesenchymal Transition in HGSOC

The SOX18 TF has been implicated in angiogenesis and lymphangiogenesis (Duong et al., 2012; François et al., 2008; Lilly et al., 2017) but has not previously been shown to have a cell-autonomous role in HGSOC. We quantified SOX18 gene expression in 13 high-grade serous ovarian cancer cell lines and 3 immortalized FTSEC lines (Figure 4A). *SOX18* transcript was overexpressed in HGSOC cell lines compared to normal FTSECs (FC = 18.5), indicating that the elevated expression of SOX18 seen in primary tumors (Figure 3K) is driven, at least in part, by endogenous tumor epithelial cell expression.

To model the role of SOX18 overexpression in HGSOC development, we ectopically overexpressed SOX18 in a *TERT*-immortalized FTSEC line (FT282) stably expressing mutant p53 (Figures 4B, 4C, and S4A). Single-cell RNA-seq analysis was performed to identify transcriptomic changes associated with SOX18 overexpression compared to controls. Graph-based clustering analysis identified 5 main clusters (Figure S4B); clusters 1 and 3 were enriched for SOX18-overexpressing cells, whereas cluster 2 was enriched for control cells (Figure 4D). Many of the differentially genes in clusters 1–3 were associated with the extracellular matrix and the epithelial-to-mesenchymal transition (EMT), including integrin subunit beta-4 (*ITGB4*) and fibronectin 1 (*FN1*). The collection of 263 genes significantly upregulated in SOX18-overexpressing cells (FC > 1.2, $p < 0.001$) was associated with extracellular matrix organization ($p_{adj} = 6.6 \times 10^{-12}$), non-integrin-membrane-ECM interactions ($p_{adj} = 4.0 \times 10^{-10}$), and wound healing ($p_{adj} = 4.7 \times 10^{-8}$), indicating epithelial differentiation was disrupted following SOX18 overexpression (Figure 4E). Consistent with this, SOX18-overexpressing FT282 cells adopted a more mesenchymal cellular morphology and exhibited longer population doubling times compared to controls (Figures 4F and 4G). OSECs overexpressing SOX18 did not show the same morphological change, and population doubling times were unaffected (Figures S4C–S4F). Mechanical phenotypes of the cells were evaluated using a parallel microfiltration assay, in which more deformable cells pass more readily through 10-μm pores in a polycarbonate membrane in response to applied pressure (Qi et al., 2015). SOX18-overexpressing FTSECs were significantly more deformable than parental or control cells (Figure 4H), consistent with observations that ectopic expression of key EMT transcription factors, SNAI1, SNAI2 or ZEB1, increases the deformability of EOC cells *in vitro* ($p < 0.01$, unpaired Student's t test) (Qi et al., 2015). Indeed, ectopic SOX18 expression in FTSECs induced upregulated expression of mesenchymal markers (*CDH2*, *PRRX1*, *SNAI1*, *SNAI2*, *TWIST1*, *VIM*, and *ZEB1*) and downregulated expression of epithelial marker *CDH1* (Figure 4I), measured using qRT-PCR. In contrast, EMT gene expression was not affected in OSECs overexpressing SOX18 (Figure S4). In FTSECs, dysregulation of EMT markers was partially rescued by *PRRX1* depletion, indicating this factor is in part responsible for the observed EMT, but other factors also likely contribute (Figure 4J).

Identifying Drivers of Transcriptional Reprogramming in the Development of HGSOC

To identify additional TFs that may drive transcriptional deregulation during HGSOC development from FTSECs and/or OSECs, we performed a targeted analysis of genes associated with DNA binding, TF activity, and chromatin remodeling. For nine of the most overexpressed transcriptional regulators in HGSOCs (Figure 5A), high-quality ChIP-seq data were available from http://cistrome.org. We quantified how many of the most differentially expressed genes in HGSOCs were located near to (within 50 kbp) a factor-specific peak compared to matched random peaks (see STAR Methods). Factor-specific peaks for regulators including SPI1, CTCFL, NFE2, ASCL2, and GRHL2 were more numerous in the vicinity of HGSOC DEGs ($p < 10^{-30}$ for a comparison to FTSECs and $p < 10^{-50}$ for a comparison to OSECs) than randomly generated matched sets of background peaks (100 iterations, $p < 0.01$; Figure 5B,C; genes located close to factor peaks are listed in Tables S3 and S4). Binding sites for ELF3, a factor highly expressed in both FTSECs and HGSOCs but lowly expressed in OSECs, were specifically enriched near to genes differentially expressed between FTSECs and HGSOCs ($p = 0.009$; Figure 5B), with no evidence of enrichment in the set of genes differentially expressed in the development of HGSOC from OSECs ($p = 1$; Figure 5C). Conversely, EHF-binding sites were associated with DEGs in a comparison of HGSOCs to OSECs but not FTSECs ($p = 0.009$ and $p = 0.06$, respectively). Notably, both ELF3 and EHF genes are proximal to SEs in HGSOC (Figure S5). Collectively, these factors represent drivers of transcriptional reprogramming in HGSOC, with ELF3 likely to be specific to the transformation of FTSECs, and EHF, a putative driver of HGSOC development from OSECs.

DISCUSSION

The cellular origins of high-grade serous ovarian cancer are debated. Over the last few years, the fallopian tube and, specifically the secretory epithelial cell component (FTSECs), has emerged as the most likely common origin for HGSOCs; but the existing data suggest there may be more than one cell of origin. In this study, we used machine learning to address the hypothesis that HGSOCs have dualistic cellular origins with ovarian surface epithelial cells (OSECs), another precursor cell type. The machine learning approach has been well established as a metric for classifying tumor of unknown origin, based on

(C, F, and I) H3K27ac ChIP-seq data were integrated with RNA-seq data for each tissue type. Average gene expression in 114 OSECs, 74 FTSECs, and 394 HGSOCs is shown for regions centered on cell-type-specific SEs; (D, G, and J) representative loci displaying tissue-specific SE deposition for each tissue type; (E, H, and K) boxplots illustrating differential gene expression between tissue types for candidate, cell-type-specific *cis*-regulated genes. The associated gene consistently displays higher expression in the SE-positive tissue type.

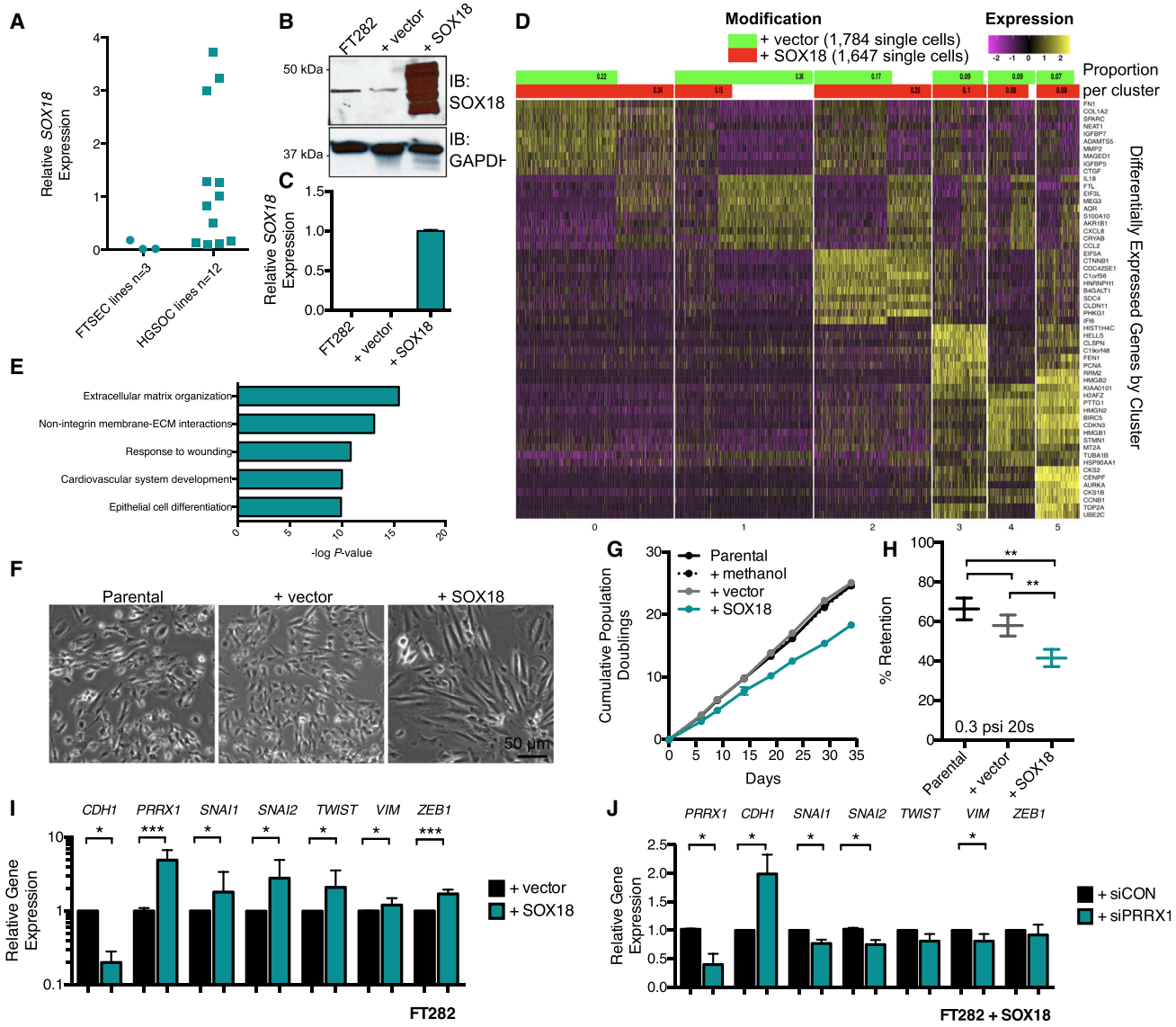


Figure 4. SOX18 Overexpression Promotes EMT

(A) SOX18 expression in normal FTSEC (n = 3) and HGSOC (n = 12) cell lines.

(B and C) Ectopic overexpression of SOX18 in FTSECs confirmed by western blotting (B) and qRT-PCR (C).

(D) Single-cell gene expression analysis in vector-transfected and SOX18-overexpressing cells. Proportions of cells represented in each cluster are represented by the green and red bars, above a heatmap showing the genes that define each cluster.

(E) Pathway analysis of genes highly expressed in SOX18-overexpressing cells.

(F) Phase-contrast images of control and SOX18-overexpressing cells.

(G) Growth curve assay showing mean cumulative population doublings ± SD (technical triplicate samples, growth curve representative of three independent experiments).

(H) Deformability measurements for cells obtained using the parallel microfiltration (PMF) method, with a 10 µm pore membrane and 2.1 kPa pressure applied for 20 s. All data points represent mean ± SD of 3 independent experiments.

(I) qRT-PCR analysis of EMT genes in SOX18-overexpressing and vector-transfected cells.

(J) EMT gene expression in SOX18-overexpressing cells with *PRRX1* or control siRNA treatment.

Data points represent mean ± SD of at least 3 independent experiments performed with technical triplicate measurements. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$; two-tailed paired Student's t test.

"bulk" molecular profiling (Flynn et al., 2018; Moran et al., 2016; Staub et al., 2010; Søndergaard et al., 2017).

The evidence for FTSECs as the major cell of origin of HGSOC is substantial: early-stage lesions in the fallopian tube, particularly in the fallopian tube fimbriae, express secretory cell lineage markers and harbor the same *TP53* mutations as metastatic tumors (Callahan et al., 2007; Gilks et al., 2015; Kindelberger et al., 2007; Kuhn et al., 2012; Paley et al., 2001). Recent genomic

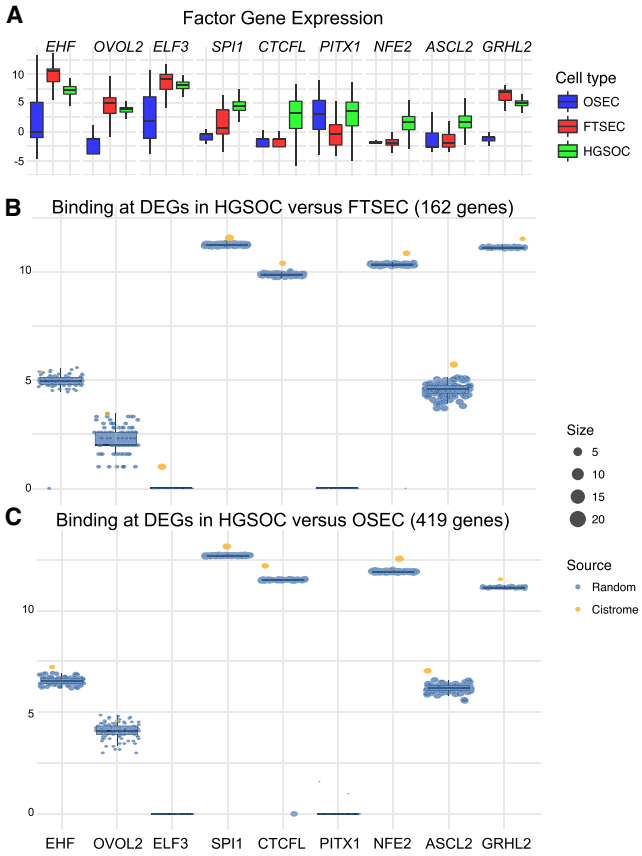


Figure 5. Transcriptional Regulators Implicated in HGSOC Development

(A) Transcriptional regulators highly expressed in HGSOCs.

(B and C) The number of differentially expressed genes in HGSOCs compared to FTSECs (B) and OSECs (C) that localize with ChIP-seq peaks for each factor (orange) and factor-specific matched random peaks (blue). *p < 0.05, **p < 0.01.

analyses have identified HGSOC-specific copy number alterations and somatic mutations in serous tubal intraepithelial carcinomas (Labidi-Galy et al., 2017) and find no significant differences in the molecular profiles of HGSOCs associated with STICs and those without (Ducie et al., 2017). *In vitro* and *in vivo* modeling studies also support FTSECs as a major cell of origin for HGSOC (Karst et al., 2011; Perets et al., 2013; Zhai et al., 2017), and salpingectomy (surgical removal of fallopian tubes but not the ovaries) can reduce the risk of ovarian cancer by around 35% or more (Falconer et al., 2015).

Historically, OSECs were thought to be the precursor cell type for HGSOC, and there remains significant, and often overlooked, evidence supporting this hypothesis. First, OSECs can express many prominent HGSOC markers, including PAX8 (Adler et al., 2015; Park et al., 2018). Second, OSECs from women at high risk of ovarian cancer are more committed to an epithelial phenotype and in culture maintain expression of CA125 longer than OSECs from non-high-risk women (Dyck et al., 1996). Third, occult cancers have been detected in the ovaries of women undergoing prophylactic risk reducing oophorectomy and can occur without evidence of lesions in the fallopian tube (Powell et al., 2005). Fourth, a recent detailed analysis of almost 60 ovaries documented evidence for metaplasia of ovarian epithelium to a Müllerian phenotype, suggesting that adoption of a more fallopian-like morphology may be an early step in the neoplastic transformation of OSECs (Park et al., 2018).

In the current study, we found greater similarities in transcriptomic signatures between FTSECs and a cohort of almost 400 HGSOCs, supporting FTSECs as the major precursor cell type. Primary HGSOCs can be sub-stratified into 4 different molecular groups based on mRNA expression profiles, and high FTSEC scores were enriched in mesenchymal-type HGSOCs, consistent with previous observations indicating that fallopian-like HGSOCs are associated with poorer clinical outcomes (Merritt et al., 2013). A small proportion of HGSOCs had transcriptional signatures that were more similar to OSECs, and these OSEC-like tumors tended to be of the proliferative molecular subgroup, which have better outcomes. Taken together, these data indicate that different molecular subgroups of HGSOC may derive from different precursor cells and suggest the cell of origin may influence patient outcomes.

Little is known about the key transcription factors (TFs) driving oncogenesis in HGSOC. The most significant TF identified to date is PAX8, which is highly expressed in FTSECs and HGSOCs, moderately expressed in some OSECs, and is functionally involved in disease development (Adler et al., 2015, 2017; Cheung et al., 2011; Elias et al., 2016; Kar et al., 2017). By analyzing epigenomic landscapes annotated from H3K27ac ChIP-seq data, we found that the *PAX8* gene locus is marked by a strong SE in both FTSECs and HGSOCs. We also identified SOX18 as one of the few TFs that coincided with a tumor-specific SE, and functional studies found SOX18 to be a regulator of EMT in HGSOC development from FTSECs but not OSECs. SOX18 has established roles in tumor-induced angiogenesis and lymphangiogenesis and is known to be aberrantly expressed in melanoma and gastric, breast, lung, and pancreatic cancers (Duong et al., 2012; Eom et al., 2012; Pula et al., 2013; Saitoh and Katoh, 2002; Zhang et al., 2016). In gastric cancer, SOX18 expression is correlated with increased lymph node metastasis and worse overall survival (Eom et al., 2012). Additionally, SOX18 levels in peripheral blood samples of gastric cancer patients are significantly increased compared to healthy controls, suggesting the feasibility of clinically assessing SOX18 levels in patients with cancer. The role of SOX18 in ovarian cancer is not well defined. One study examined SOX18 protein expression in a small cohort of 85 patients and found high SOX18 expression was associated with advanced stage and worse disease specific survival (Pula et al., 2014). Our finding that SOX18 induces EMT in FTSECs *in vitro* may indicate that SOX18 plays a role in establishing the mesenchymal signature enriched in FTSEC-like HGSOCs.

In conclusion, this study represents a significant advance on previous studies in both scale and scope of the molecular profiling of putative precursor cell types for high-grade serous ovarian cancer (Merritt et al., 2013). This represents, considerably, the largest study to profile the transcriptomes of FTSECs and OSECs (188 samples from 135 different individuals in total), which enabled us to derive robust signatures with which to study

the relationships between these putative cells of origin and the different molecular subtypes of HGSOC. As a result, we have been able to provide evidence that both FTSECs and OSECs are likely precursors of this disease, and we note that our conclusions are consistent with those observed in a new mouse model (Zhang et al., 2018). Moreover, we provide evidence for the role of SOX18 and other transcription factors in the development of HGSOC, which may represent candidate clinical biomarkers and potential therapeutic targets for this disease.

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
- METHOD DETAILS
  - Sample collection, RNA extraction and RNA sequencing
  - Tissue ChIP-seq in HGSOC specimens
  - Cell culture and SOX18 overexpression
  - Single cell RNA-seq data generation
  - Growth curves
  - PMF assay
  - siRNA interference
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - RNA-seq data processing and QC
  - Differential gene expression analyses
  - Machine learning analyses
  - ChIP-seq data analysis
  - Single cell RNA-seq data analyses
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2019.10.122.

ACKNOWLEDGMENTS

The results shown here are, in part, based upon data generated by the TCGA Research Network: https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga. Some of the normal tissue specimens were collected as part of the USC Jean Richardson Gynecologic Tissue and Fluid Repository, which is supported by a grant from the USC Department of Obstetrics & Gynecology and the NCT Cancer Center Shared Grant award P30 CA014089 (to the Norris Comprehensive Cancer Center). K.L. is supported, in part, by a K99/R00 Pathway to Independence Award from the NIH (R00CA184415), institutional support from the Samuel Oschin Comprehensive Cancer Institute at Cedars-Sinai Medical Center, and a Career Development Award from The Tower Cancer Research Foundation. M.A.S.F. is supported by grant 2017/08211-1 from Sao Paulo Research Foundation (FAPESP). H.N. is also supported by an institutional grant (Henry Ford Hospital). N.K.G. is supported by Farber Family Funds. This work was supported, in part, by the Ovarian Cancer Research Fund Alliance Program Project Development Grant (373356): Co-Evolution of Epithelial Ovarian Cancer and Tumor Stroma. Additional support for this work came from NIH/NCI grants 1R01CA211707 and 1R01CA207456 and OCRF award 258807, as well as the Department of Defense Ovarian Cancer Research Program Pilot Award (A.C.R., K.L., and B.Y.K.).

AUTHOR CONTRIBUTIONS

Study Concept, Design, and Manuscript Preparation, K.L., S.A.G., and H.N.; Data Generation, K.L., A.Y.L. J.M.L., X.L., A.F., N.K.G., and J.-H.S.; Data Analysis, M.A.S.F., F.S., and H.Q.D.; Sample/Cell Line Provision, Y.G.L., R.D., and B.Y.K.; Pathology Review, P.M.-F.; Technical/Analytic Expertise, K.C.V., S.C., A.C.R., D.J.H., and M.L.F. All authors contributed to the final draft of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 20, 2018
Revised: August 6, 2019
Accepted: October 29, 2019
Published: December 10, 2019

REFERENCES

Adler, E., Mhawech-Fauceglia, P., Gayther, S.A., and Lawrenson, K. (2015). PAX8 expression in ovarian surface epithelial cells. Hum. Pathol. *46*, 948–956.

Adler, E.K., Corona, R.I., Lee, J.M., Rodriguez-Malave, N., Mhawech-Fauceglia, P., Sowter, H., Hazelett, D.J., Lawrenson, K., and Gayther, S.A. (2017). The PAX8 cistrome in epithelial ovarian cancer. Oncotarget *8*, 108316–108332.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Callahan, M.J., Crum, C.P., Medeiros, F., Kindelberger, D.W., Elvin, J.A., Garber, J.E., Feltmate, C.M., Berkowitz, R.S., and Muto, M.G. (2007). Primary fallopian tube malignancies in BRCA-positive women undergoing surgery for ovarian cancer risk reduction. J. Clin. Oncol. *25*, 3985–3990.

Cancer Genome Atlas Research Network. (2011). Integrated genomic analyses of ovarian carcinoma. Nature *474*, 609–615.

Cheung, H.W., Cowley, G.S., Weir, B.A., Boehm, J.S., Rusin, S., Scott, J.A., East, A., Ali, L.D., Lizotte, P.H., Wong, T.C., et al. (2011). Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. Proc. Natl. Acad. Sci. USA *108*, 12372–12377.

Coetzee, S.G., Shen, H.C., Hazelett, D.J., Lawrenson, K., Kuchenbaecker, K., Tyrer, J., Rhie, S.K., Levanon, K., Karst, A., Drapkin, R., et al.; Ovarian Cancer Association Consortium, The Consortium of Investigators of Modifiers of BRCA1/2; Ovarian Cancer Association Consortium The Consortium of Investigators of Modifiers of BRCA1/2 (2015). Cell-type-specific enrichment of risk-associated regulatory elements at ovarian cancer susceptibility loci. Hum. Mol. Genet. *24*, 3595–3607.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

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.

Duong, T., Proulx, S.T., Luciani, P., Leroux, J.-C., Detmar, M., Koopman, P., and Francois, M. (2012). Genetic ablation of SOX18 function suppresses tumor lymphangiogenesis and metastasis of melanoma in mice. Cancer Res. *72*, 3105–3114.

Dyck, H.G., Hamilton, T.C., Godwin, A.K., Lynch, H.T., Maines-Bandiera, S., and Auersperg, N. (1996). Autonomy of the epithelial phenotype in human ovarian surface epithelium: changes with neoplastic progression and with a family history of ovarian cancer. Int. J. Cancer *69*, 429–436.

Eckert, M.A., Pan, S., Hernandez, K.M., Loth, R.M., Andrade, J., Volchenboum, S.L., Faber, P., Montag, A., Lastra, R., Peter, M.E., et al. (2016). Genomics of ovarian cancer progression reveals diverse metastatic trajectories including intraepithelial metastasis to the fallopian tube. Cancer Discov. *6*, 1342–1351.

Elias, K.M., Emori, M.M., Westerling, T., Long, H., Budina-Kolomets, A., Li, F., MacDuffie, E., Davis, M.R., Holman, A., Lawney, B., et al. (2016). Epigenetic remodeling regulates transcriptional changes between ovarian cancer and benign precursors. JCI Insight *1*, e87988.

Eom, B.W., Jo, M.J., Kook, M.-C., Ryu, K.W., Choi, I.J., Nam, B.-H., Kim, Y.-W., and Lee, J.H. (2012). The lymphangiogenic factor SOX 18: a key indicator to stage gastric tumor progression. Int. J. Cancer *131*, 41–48.

Falconer, H., Yin, L., Grönberg, H., and Altman, D. (2015). Ovarian cancer risk after salpingectomy: a nationwide population-based study. J. Natl. Cancer Inst. *107*, dju410.

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. *16*, 278.

Flynn, W.F., Namburi, S., Paisie, C.A., Reddi, H.V., Li, S., Karuturi, R.K.M., and George, J. (2018). Pan-cancer machine learning predictors of tissue of origin and molecular subtype. bioRxiv. https://doi.org/10.1101/333914.

François, M., Caprini, A., Hosking, B., Orsenigo, F., Wilhelm, D., Browne, C., Paavonen, K., Karnezis, T., Shayan, R., Downes, M., et al. (2008). Sox18 induces development of the lymphatic vasculature in mice. Nature *456*, 643–647.

Gilks, C.B., Irving, J., Köbel, M., Lee, C., Singh, N., Wilkinson, N., and McCluggage, W.G. (2015). Incidental nonuterine high-grade serous carcinomas arise in the fallopian tube in most cases: further evidence for the tubal origin of high-grade serous carcinomas. Am. J. Surg. Pathol. *39*, 357–364.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Kar, S.P., Adler, E., Tyrer, J., Hazelett, D., Anton-Culver, H., Bandera, E.V., Beckmann, M.W., Berchuck, A., Bogdanova, N., Brinton, L., et al. (2017). Enrichment of putative PAX8 target genes at serous epithelial ovarian cancer susceptibility loci. Br. J. Cancer *116*, 524–535.

Karst, A.M., Levanon, K., and Drapkin, R. (2011). Modeling high-grade serous ovarian carcinogenesis from the fallopian tube. Proc. Natl. Acad. Sci. USA *108*, 7547–7552.

Kindelberger, D.W., Lee, Y., Miron, A., Hirsch, M.S., Feltmate, C., Medeiros, F., Callahan, M.J., Garner, E.O., Gordon, R.W., Birch, C., et al. (2007). Intraepithelial carcinoma of the fimbria and pelvic serous carcinoma: Evidence for a causal relationship. Am. J. Surg. Pathol. *31*, 161–169.

Kruk, P.A., and Auersperg, N. (1992). Human ovarian surface epithelial cells are capable of physically restructuring extracellular matrix. Am. J. Obstet. Gynecol. *167*, 1437–1443.

Kuhn, E., Kurman, R.J., Vang, R., Sehdev, A.S., Han, G., Soslow, R., Wang, T.-L., and Shih, IeM. (2012). TP53 mutations in serous tubal intraepithelial carcinoma and concurrent pelvic high-grade serous carcinoma–evidence supporting the clonal relationship of the two lesions. J. Pathol. *226*, 421–426.

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.

Laury, A.R., Perets, R., Piao, H., Krane, J.F., Barletta, J.A., French, C., Chirieac, L.R., Lis, R., Loda, M., Hornick, J.L., et al. (2011). A comprehensive analysis of PAX8 expression in human epithelial tumors. Am. J. Surg. Pathol. *35*, 816–826.

Lawrenson, K., Benjamin, E., Turmaine, M., Jacobs, I., Gayther, S., and Dafou, D. (2009). In vitro three-dimensional modelling of human ovarian surface epithelial cells. Cell Prolif. *42*, 385–393.

Lee, J.M., Mhawech-Fauceglia, P., Lee, N., Parsanian, L.C., Lin, Y.G., Gayther, S.A., and Lawrenson, K. (2013). A three-dimensional microenvironment alters protein expression and chemosensitivity of epithelial ovarian cancer cells in vitro. Lab. Invest. *93*, 528–542.

Leek, J.T., Johnson, W.E., Parker, H.S., Fertig, E.J., Jaffe, A.E., Storey, J.D., Zhang, Y., and Torres, L.C. (2019). sva: Surrogate Variable Analysis. R package version 3.34.0.

Leeper, K., Garcia, R., Swisher, E., Goff, B., Greer, B., and Paley, P. (2002). Pathologic findings in prophylactic oophorectomy specimens in high-risk women. Gynecol. Oncol. *87*, 52–56.

Levanon, K., Ng, V., Piao, H.Y., Zhang, Y., Chang, M.C., Roh, M.H., Kindelberger, D.W., Hirsch, M.S., Crum, C.P., Marto, J.A., and Drapkin, R. (2010). Primary ex vivo cultures of human fallopian tube epithelium as a model for serous ovarian carcinogenesis. Oncogene *29*, 1103–1113.

Li, N.F., Broad, S., Lu, Y.J., Yang, J.S., Watson, R., Hagemann, T., Wilbanks, G., Jacobs, I., Balkwill, F., Dafou, D., and Gayther, S.A. (2007). Human ovarian surface epithelial cells immortalized with hTERT maintain functional pRb and p53 expression. Cell Prolif. *40*, 780–794.

Lilly, A.J., Lacaud, G., and Kouskoff, V. (2017). SOXF transcription factors in cardiovascular development. Semin. Cell Dev. Biol. *63*, 50–57.

Manek, R., Pakzamir, E., Mhawech-Fauceglia, P., Pejovic, T., Sowter, H., Gayther, S.A., and Lawrenson, K. (2016). Targeting Src in endometriosis-associated ovarian cancer. Oncogenesis *5*, e251.

Medeiros, F., Muto, M.G., Lee, Y., Elvin, J.A., Callahan, M.J., Feltmate, C., Garber, J.E., Cramer, D.W., and Crum, C.P. (2006). The tubal fimbria is a preferred site for early adenocarcinoma in women with familial ovarian cancer syndrome. Am. J. Surg. Pathol. *30*, 230–236.

Merritt, M.A., Bentink, S., Schwede, M., Iwanicki, M.P., Quackenbush, J., Woo, T., Agoston, E.S., Reinhardt, F., Crum, C.P., Berkowitz, R.S., et al. (2013). Gene expression signature of normal cell-of-origin predicts ovarian tumor outcomes. PLoS One *8*, e80314.

Mhawech-Fauceglia, P., Wang, D., Samrao, D., Godoy, H., Ough, F., Liu, S., Pejovic, T., and Lele, S. (2012). Pair Box 8 (PAX8) protein expression in high grade, late stage (stages III and IV) ovarian serous carcinoma. Gynecol. Oncol. *127*, 198–201.

Moran, S., Martínez-Cardús, A., Sayols, S., Musulén, E., Balañá, C., Estival-Gonzalez, A., Moutinho, C., Heyn, H., Diaz-Lagares, A., de Moura, M.C., et al. (2016). Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. Lancet Oncol. *17*, 1386–1395.

Paley, P.J., Swisher, E.M., Garcia, R.L., Agoff, S.N., Greer, B.E., Peters, K.L., and Goff, B.A. (2001). Occult cancer of the fallopian tube in BRCA-1 germline mutation carriers at prophylactic oophorectomy: a case for recommending hysterectomy at surgical prophylaxis. Gynecol. Oncol. *80*, 176–180.

Park, K.J., Patel, P., Linkov, I., Jotwani, A., Kauff, N., and Pike, M.C. (2018). Observations on the origin of ovarian cortical inclusion cysts in women undergoing risk-reducing salpingo-oophorectomy. Histopathology *72*, 766–776.

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 Brca;Tp53;Pten models. Cancer Cell *24*, 751–765.

Piek, J.M., van Diest, P.J., Zweemer, R.P., Jansen, J.W., Poort-Keesom, R.J., Menko, F.H., Gille, J.J., Jongsma, A.P., Pals, G., Kenemans, P., and Verheijen, R.H. (2001). Dysplastic changes in prophylactically removed Fallopian tubes of women predisposed to developing ovarian cancer. J. Pathol. *195*, 451–456.

Pomerantz, M.M., Li, F., Takeda, D.Y., Lenci, R., Chonkar, A., Chabot, M., Cejas, P., Vazquez, F., Cook, J., Shivdasani, R.A., et al. (2015). The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. Nat. Genet. *47*, 1346–1351.

Pothuri, B., Leitao, M.M., Levine, D.A., Viale, A., Olshen, A.B., Arroyo, C., Bogomolniy, F., Olvera, N., Lin, O., Soslow, R.A., et al. (2010). Genetic analysis of the early natural history of epithelial ovarian carcinoma. PLoS One *5*, e10358.

Powell, C.B., Kenley, E., Chen, L.-M., Crawford, B., McLennan, J., Zaloudek, C., Komaromy, M., Beattie, M., and Ziegler, J. (2005). Risk-reducing salpingo-

oophorectomy in BRCA mutation carriers: role of serial sectioning in the detection of occult malignancy. J. Clin. Oncol. *23*, 127–132.

Pula, B., Olbromski, M., Wojnar, A., Gomulkiewicz, A., Witkiewicz, W., Ugorski, M., Dziegiel, P., and Podhorska-Okolow, M. (2013). Impact of SOX18 expression in cancer cells and vessels on the outcome of invasive ductal breast carcinoma. Cell Oncol. (Dordr.) *36*, 469–483.

Pula, B., Kobierzycki, C., Solinski, D., Olbromski, M., Nowak-Markwitz, E., Spaczynski, M., Kedzia, W., Zabel, M., and Dziegiel, P. (2014). SOX18 expression predicts response to platinum-based chemotherapy in ovarian cancer. Anticancer Res. *34*, 4029–4037.

Qi, D., Kaur Gill, N., Santiskulvong, C., Sifuentes, J., Dorigo, O., Rao, J., Taylor-Harding, B., Ruprecht Wiedemeyer, W., and Rowat, A.C. (2015). Screening cell mechanotype by parallel microfiltration. Sci. Rep. *5*, 17595.

Risso, D., Schwartz, K., Sherlock, G., and Dudoit, S. (2011). GC-content normalization for RNA-Seq data. BMC Bioinformatics *12*, 480.

Saitoh, T., and Katoh, M. (2002). Expression of human SOX18 in normal tissues and tumors. Int. J. Mol. Med. *10*, 339–344.

Salazar, H., Godwin, A.K., Daly, M.B., Laub, P.B., Hogan, W.M., Rosenblum, N., Boente, M.P., Lynch, H.T., and Hamilton, T.C. (1996). Microscopic benign and invasive malignant neoplasms and a cancer-prone phenotype in prophylactic oophorectomies. J. Natl. Cancer Inst. *88*, 1810–1820.

Sergushichev, A. (2016). An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. bioRxiv. https://doi.org/10.1101/060012.

Sokolov, A., Paull, E.O., and Stuart, J.M. (2016a). One-class detection of cell states in tumor subtypes. Pac. Symp. Biocomput. *21*, 405–416.

Sokolov, A., Carlin, D.E., Paull, E.O., Baertsch, R., and Stuart, J.M. (2016b). Pathway-Based Genomics Prediction using Generalized Elastic Net. PLoS Comput. Biol. *12*, e1004790.

Søndergaard, D., Nielsen, S., Pedersen, C.N.S., and Besenbacher, S. (2017). Prediction of primary tumors in cancers of unknown primary. J. Integr. Bioinform. *14*, 20170013.

Staub, E., Buhr, H.J., and Gröne, J. (2010). Predicting the site of origin of tumors by a gene expression signature derived from normal tissues. Oncogene *29*, 4485–4492.

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.; Australian Ovarian Cancer Study Group (2008). Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. Clin. Cancer Res. *14*, 5198–5208.

Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. Bioinformatics *28*, 2184–2185.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Zappia, L., and Oshlack, A. (2018). Clustering trees: a visualization for evaluating clusterings at multiple resolutions. Gigascience *7*, giy083.

Zhai, Y., Wu, R., Kuick, R., Sessine, M.S., Schulman, S., Green, M., Fearon, E.R., and Cho, K.R. (2017). High-grade serous carcinomas arise in the mouse oviduct via defects linked to the human disease. J. Pathol. *243*, 16–25.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhang, J., Ma, Y., Wang, S., Chen, F., and Gu, Y. (2016). Suppression of SOX18 by siRNA inhibits cell growth and invasion of breast cancer cells. Oncol. Rep. *35*, 3721–3727.

Zhang, S., Zhang, T., Dolgalev, I., Ran, H., Levine, D.A., and Neel, B.G. (2018). Both Fallopian Tube and Ovarian Surface Epithelium Can Act as Cell-of-Origin for High Grade Serous Ovarian Carcinoma. bioRxiv. https://doi.org/10.1101/481200.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-SOX18	Santa Cruz Biotechnology	sc-166025
Anti-GAPDH	Santa Cruz Biotechnology	sc-47724
Anti-actin	Santa Cruz Biotechnology	sc-47778
Biological Samples		
Primary ovarian surface epithelial cell (OSEC) and fallopian tube secretory epithelial cell (FTSEC) cultures	Patients (this paper)	NA
Critical Commercial Assays		
Chromium Single Cell 3′ Library & Gel Bead Kit v2	10X Genomics	PN-120237
Deposited Data		
Raw feature counts matrix, pre-processed data, superenhancer peaks	This paper	https://lawrenson-lab.github.io/OvarianRNASeq/index.html
Transcription factor peaks (bed files)	This paper, processed from http://cistrome.org	NA
Experimental Models: Cell Lines		
FT282	Ronny Drapkin	NA
IOSE4	This laboratory	NA
Oligonucleotides		
SMARTpool siRNA PRRX1	Dharmacon	L-012402-00-0005
Recombinant DNA		
SOX18 overexpression vector	Genecopoeia	EX-W0082-Lv125
Software and Algorithms		
STAR	Dobin et al., 2013	https://github.com/alexdobin/STAR
RSeQC	Wang et al., 2012	https://github.com/dnanexus/rseqc
EDASeq	Risso et al., 2011	https://bioconductor.org/packages/devel/bioc/html/EDASeq.html
Sva (combat)	Leek et al., 2019	https://bioconductor.org/packages/release/bioc/html/sva.html
gelnet	Sokolov et al., 2016b	https://github.com/ArtemSokolov/gelnet
fgsea	Sergushichev, 2016	http://bioconductor.org/packages/release/bioc/html/fgsea.html
macs2	Zhang et al., 2008	https://pypi.org/pypi/MACS2
homer	Heinz et al., 2010	http://homer.ucsd.edu/homer/
Cell Ranger	NA	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/installation
Seurat	Butler et al., 2018	https://github.com/satijalab/seurat

LEAD CONTACT AND MATERIALS AVAILABILITY

Please contact Simon A. Gayther, Center for Bioinformatics and Functional Genomics, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, California, USA; email: simon.gayther@cshs.org for further information or requests for resources and reagents. SOX18 overexpressing fallopian and ovarian cell models are available on request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

114 ovarian and 74 fallopian tube epithelial specimens were collected for this study, from women undergoing gynecologic surgeries at University College Hospital (London, UK), LAC + USC Medical Center (Los Angeles, CA, USA) and Oregon Health & Science University (Portland, OR, USA). All were collected with informed patient consent and Institutional Review Board approval. Patient information, including age, diagnosis, ace, ethnicity, and histology of any cancer diagnosed at the time of surgery are provided in Table S1 (where available). We did not perform a sample size estimation to design this study. Primary cultures were maintained in culture for a short time and were not authenticated as reference genotypes do not exist for these new culture isolates.

METHOD DETAILS

Sample collection, RNA extraction and RNA sequencing

OSECs and FTSECs were harvested from ovaries and fallopian tubes of women diagnosed with ovarian, uterine or cervical cancer. Tissues were grossly and histologically normal. Short-term cultures were established as previously described (Karst et al., 2011; Lawrenson et al., 2009). Briefly, OSECs were harvested using a cytobrush and cultured in NOSE-CM media containing 15% fetal bovine serum (FBS, Hyclone), 34 $\mu g\ ml^{-1}$ bovine pituitary extract, 10 $ng\ ml^{-1}$ epidermal growth factor (Life Technologies), 5 $\mu g\ ml^{-1}$ insulin and 500 $ng\ ml^{-1}$ hydrocortisone (Sigma-Aldrich). FTSECs were harvested by Pronase/DNase I digestion (Roche and Sigma-Aldrich, respectively) for 48-72 hours at 4°C and cultured on collagen I (Sigma-Aldrich) using DMEM/F12 base media supplemented with 2% Ultroser G (Pall Corporation). This approach is known to enrich for secretory epithelial cells over time (Levanon et al., 2010), consistent with this, our FTSEC cultures express high levels of PAX8 (Figure S1B). Five samples were sequenced twice to ensure replication (Figure S1A).

At ~80% confluency, cells were lysed using the QIAzol reagent and RNA extracted using the RNeasy Mini kit (both QIAgen). RNA sequencing was performed at the University of Southern California Epigenome Core Facility.

Tissue ChIP-seq in HGSOC specimens

Tissue ChIP-seq was performed based on the methods described in Pomerantz et al. (2015). One 3 mm core was isolated from epithelial-rich portions of four high-grade serous ovarian cancers, and pulverized using the Covaris CryoPrep system (Covaris, Woburn, MA), set to an intensity of 4. Tissues were fixed using 1% formaldehyde (Thermo fisher, Waltham, MA) for 10 minutes at room temperature. Fixation was quenched with 125 mM glycine and samples were rinsed with cold PBS before a 10 minute lysis in a buffer containing 50 mM Tris, 10mM EDTA, 1% SDS with protease inhibitor). Chromatin was sheared to 300–500 base pairs and 5 volumes dilution buffer (1% Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris HCl pH 8.1) added. Each sample was incubated with 1 μg H3K27ac antibody (DiAGenode, C15410196, Denville, NJ) coupled with protein A and protein G beads (Life Technologies, Carlsbad, CA) at 4°C overnight. Immunoprecipitated chromatin was washed with RIPA buffer (0.05M HEPES pH 7.6, 1 mM EDTA, 0.7% Na Deoxycholate, 1% NP-40, 0.5M LiCl) five times and rinsed with TE buffer (pH 8.0) once. The sample was resuspended in elution buffer (50mM Tris, 10mM EDTA, 1% SDS), treated with RNase for 30 minutes at 37°C, and incubated with proteinase K overnight at 65°C. Sample DNA and 1% input were extracted, and sequencing libraries prepared using the ThruPLEX-FD Prep Kit (Rubicon Genomics, Ann Arbor, MI). Libraries were sequenced using 75-base pair single reads on the Illumina platform (Illumina, San Diego, CA) at the Dana-Farber Cancer Institute.

Cell culture and SOX18 overexpression

Human fallopian tube cell lines FT246, FT282, and FT318 were grown in DMEM/F12 with 10% FBS. SOX18 expression was surveyed in the following cell lines: CaOV3, COV318, EFO21, Kuramochi, FUOV1, OAW28, OV177, OVSAHO, TykNu, UWB1.289; details of cell culture media can be found in Lee et al. (2013) and Manek et al. (2016). A SOX18 overexpression vector and vector control were purchased from Genecopoeia (Rockville, MD), and DNA was extracted using the Maxiprep kit (QIAgen, Hilden, Germany). Lentiviruses were generated in HEK293T cells by transient transfection with BioT transfection reagent (Bioland Scientific, Paramount, CA) according to manufacturer's instructions. Lentivirus-containing medium was collected 48 hours after transfection and filtered through a 0.22μM filter (Millipore, Burlington, MA). The lentivirus-containing media was added to culture medium with 8μg/mL Polybrene transfection reagent (Sigma-Aldrich, St. Louis, MO). Selection was performed with 1000 μg/mL puromycin diluted in absolute methanol, and SOX18 and vector control transduced cell lines were then maintained with 1000 μg/mL puromycin.

Single cell RNA-seq data generation

FT282 cells with empty vector control or stable SOX18 overexpression were processed into single cells by trypsinization. Single cell RNA-seq libraries were made using the Chromium Single Cell 3′ Library & Gel Bead Kit v2 (10X Genomics, Catalogue number PN-120237) according to manufacturer's instructions. A total of 2,000 cells were targeted for recovery. The scRNA-Seq libraries were pooled and sequenced with paired-end 150 bp reads on the HiSeq 4000 platform at Fulgent Genetics (http://fulgentgenetics.com).

Growth curves

Cells were plated at 50,000 cells per well of a 6-well plate, in triplicate. An additional triplicate of the parental cell line was treated with absolute methanol to serve as an additional control. Cells were passaged and counted every 3-4 days for > 28 days. Growth curves were performed three times, independently.

PMF assay

The PMF device is assembled using polycarbonate membrane (Isopore, Millipore) with 10 μm pore diameter. Cell suspension (350 μL) at a concentration of 0.5×10^6 cells/mL is loaded into each well. Constant air pressure of 2.1 kPa is applied for 20 s using a custom-built manometer and monitored using a pressure gauge (Noshok Inc., Berea, OH, USA). We determine % retention by collecting the sample suspension remaining in the top well and reading absorbance at 560 nm wavelength of the retained volume using a plate reader (Techan Infinite M1000, Thermo Fisher Scientific). To measure cell number and obtain size distributions, we use an automated cell counter (TC20, BioRad). All data points are obtained from 3 independent experiments with 3 replicate wells per sample. We use the Student's t test method to analyze the results and obtain p values.

siRNA interference

SMARTpool siRNAs directed against human *PRRX1* and non-targeting control were purchased from Dharmacon (Lafayette, CO). 50,000 FT282 cells stably overexpressing SOX18 were grown in 10 cm dishes. These were transfected with 50 μL of 5 μM *PRRX1* or non-targeting siRNA using DharmaFECT3 transfection reagent (Dharmacon) according to manufacturer's instructions. 6 days after transfection, RNA was harvested from transfected cells and used for RT-qPCR, performed using TaqMan probes. Each experiment was performed three times independently, with technical triplicates. Paired Student's t tests were performed to obtain p values from comparing the mean expression value from each replicate experiment.

QUANTIFICATION AND STATISTICAL ANALYSIS

RNA-seq data processing and QC

All data analysis was performed using 'R' and 'Bioconductor', and packages therein. RNaseq data for 394 HGSOC samples was obtained from The Cancer Genome Atlas (TCGA) data portal as protected data (raw sequencing, fastq files) and downloaded via CGHub's geneTorrent. Data was aligned to a reference genome (hg19) using STAR (Dobin et al., 2013) and quality control of aligned samples performed using RSeQC (Wang et al., 2012). GC bias and batch effect corrections were performed using EDASeq and 'sva' (Risso et al., 2011). To adjust for batch effects we used an empirical Bayes framework (comBat), available in 'sva'. Genes absent in more than 80% of the samples were removed. Expression values correspond to the normalized adjusted values obtained from comBat.

Differential gene expression analyses

After normalization, the data matrix contained 21,071 genes. Parametric statistics (Student's t test) and supervised hierarchical clustering were performed to identify genes differentially expressed in pairwise comparisons of two groups of interest (OSEC, FTSEC and HGSOC). P values were adjusted using Benjamini–Hochberg step-up procedure.

Machine learning analyses

We applied a machine learning approach to define a probabilistic score associated to both normal cell types and infer tumor origins. A One-class classifier was selected as this method can handle non-traditional supervised scenarios where no negative class can be defined. The classifiers were implemented by the gelnet R-package version 1.2.1 (Sokolov et al., 2016b). Data were mean centered considering all samples together, then each cell type used separately to train and test the models. To train the OSEC model we considered all OSEC samples, with a coefficient for the L1-norm penalty equal to 0 and coefficient for the L2-norm penalty equal to 1 as arguments of gelnet function. The training optimization is terminated after the desired tolerance is achieved (default 1e-5). We then evaluated the model performance through leave-one-out procedure where the left-out OSEC sample was mixed into FTSEC sample background. The accuracy was evaluated via the Area Under the ROC curve method, with 99% of OSEC samples correctly predicted, on average, and 97% of FTSECs correctly predicted. We then used the models to prediction cellular origins of 394 HGSOCs from TCGA. We took advantage of the fast gene set enrichment analysis (fgsea, version 1.2.1, http://bioconductor.org/packages/release/bioc/html/fgsea.html) (Sergushichev, 2016) method to evaluate enrichment of clinical attributes across the tumor OCLR scores from both FTSEC and OSEC models. We applied the fgsea function with the parameter nperm equal to 10,000.

ChIP-seq data analysis

The AQUAS pipeline (https://github.com/kundajelab/chipseq_pipeline) was used to processed ChIP-seq data. Reads were aligned to the reference human genome (hg19), filtered by read quality and duplicate reads removed. macs2 (https://pypi.org/pypi/MACS2; Zhang et al., 2008) was used for peak calling. For the cell lines, two technical replicates were generated and the final peaks were obtained using a naive overlap approach, where the peaks are included if they overlap more than 50% between the two technical replicates. We have previously described H3K27ac ChIP-seq for immortalized OSEC and FTSEC lines (Coetzee et al., 2015).

Immortalized OSECs have been previously shown to be representative of unmodified cells (Li et al., 2007). We verified that the expression profiles of immortalized FTSECs used in this study clustered with primary FTSECs (data not shown). After alignment, homer (http://homer.ucsd.edu/homer/) (Heinz et al., 2010) was used to identify super-enhancers, using a super slope parameter of 2 and a minimum distance of ten thousand. For defining a set of HGSOC SEs, we selected SEs that were called in at least two HGSOC samples. For the FTSEC set of SEs, SEs were called individually in each technical replicate, then, all the SEs that overlapped both technical replicates within the same cell line (FTSEC33 or FTSEC246) were selected to get the union set. We used a similar approach to get the union set of SEs for the OSEC cell lines.

Single cell RNA-seq data analyses

Raw reads were aligned to hg38 reference genome, UMI (unique molecular identifier) counting was done using Cell Ranger v.2.1.1 (10xGenomics) pipeline with default parameters. This yielded 1,784 cells (4,569 genes per cell) and 1,647 cells (5,199 genes per cell) for empty vector and SOX18 OE, respectively. After removing cells with high mitochondrial content (> = 20%), 1,690 and 1,543 cells were kept for downstream analysis. We used Seurat v.3.0 (Butler et al., 2018) for data integration and alignment using canonical correlation analysis (CCA) with top 12 CC dimensions as suggested by CC bicor saturation plot. Seurat graph-based clustering was then used to infer 5 clusters at resolution 0.6 after evaluating by clustree method at different resolutions from 0.2 to 1.4 (Zappia and Oshlack, 2018). Gene signatures for each subset were inferred using differentially expression analysis against the rest with MAST method (Finak et al., 2015) implemented in Seurat with default parameters. Top 10 genes for each subset were plot as heatmap using DoHeatmap function in Seurat package.

DATA AND CODE AVAILABILITY

All raw RNA sequencing data and custom code are accessible at https://lawrenson-lab.github.io/OvarianRNASeq/index.html. ChIP-seq data have been deposited into GEO (under accession number GSE121103).