HHS Public Access

Author manuscript

Gynecol Oncol. Author manuscript; available in PMC 2018 March 01.

Published in final edited form as:

Gynecol Oncol. 2017 March ; 144(3): 598–606. doi:10.1016/j.ygyno.2017.01.015.

Single cell sequencing reveals heterogeneity within ovarian cancer epithelium and cancer associated stromal cells

Boris J. Winterhoff[a,b], **Makayla Maile**[a,b], **Amit Kumar Mitra**[c], **Attila Sebe**[d], **Martina Bazzaro**[a,b], **Melissa A. Geller**[a,b], **Juan E. Abrahante**[e], **Molly Klein**[f], **Raffaele Hellweg**[a], **Sally A. Mullany**[a], **Kenneth Beckman**[g], **Jerry Daniel**[g], and **Timothy K. Starr**[a,b,*]

[a]Department of Obstetrics, Gynecology & Women's Health, University of Minnesota, Minneapolis, MN, USA

[b]Masonic Cancer Center, University of Minnesota, Minneapolis, MN, USA

[c]Department of Genetics, Cell Biology & Development, University of Minnesota, Minneapolis, MN, USA

[d]Division of Medical Biotechnology, Paul Ehrlich Institute, Langen, Germany

[e]University of Minnesota Informatics Institute, University of Minnesota, Minneapolis, MN, USA

[f]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, Minnesota, USA

[g]Genomics Center, University of Minnesota, Minneapolis, MN, USA

Abstract

Objectives—The purpose of this study was to determine the level of heterogeneity in high grade serous ovarian cancer (HGSOC) by analyzing RNA expression in single epithelial and cancer associated stromal cells. In addition, we explored the possibility of identifying subgroups based on pathway activation and pre-defined signatures from cancer stem cells and chemo-resistant cells.

Methods—A fresh, HGSOC tumor specimen derived from ovary was enzymatically digested and depleted of immune infiltrating cells. RNA sequencing was performed on 92 single cells and 66 of these single cell datasets passed quality control checks. Sequences were analyzed using multiple bioinformatics tools, including clustering, principle components analysis, and geneset enrichment

[*]**Corresponding author**: Timothy K. Starr, MMC 395, 420 Delaware St SE, Minneapolis, MN, 55455. Phone: 612-625-4425, Fax: 612-626-0665, star0044@umn.edu.

Author Contributions

The author contributions to this work are as follows: BJW, MAG, and SAM conducted patient selection and sample collection. MM and TKS performed single cell isolation. AKM, AS, JEA, TKS and RH performed bioinformatic and statistical analysis. MK analyzed pathology slides. KB and JD performed Fluidigm C1 and Illumina HiSeq analyses. MB, BJW and TKS wrote the manuscript.

Conflict of interest Statement

The authors declare that there are no conflicts of interest.

analysis to identify subgroups and activated pathways. Immunohistochemistry for ovarian cancer, stem cell and stromal markers was performed on adjacent tumor sections.

Results—Analysis of the gene expression patterns identified two major subsets of cells characterized by epithelial and stromal gene expression patterns. The epithelial group was characterized by proliferative genes including genes associated with oxidative phosphorylation and MYC activity, while the stromal group was characterized by increased expression of extracellular matrix (ECM) genes and genes associated with epithelial-to-mesenchymal transition (EMT). Neither group expressed a signature correlating with published chemo-resistant gene signatures, but many cells, predominantly in the stromal subgroup, expressed markers associated with cancer stem cells.

Conclusions—Single cell sequencing provides a means of identifying subpopulations of cancer cells within a single patient. Single cell sequence analysis may prove to be critical for understanding the etiology, progression and drug resistance in ovarian cancer.

Keywords

Ovarian cancer; molecular subtypes; single cell sequencing and cancer stem cells

Introduction

The promise of individualized cancer therapy is predicated on the identification of remediable drug targets within a tumor. Target identification has traditionally relied on technologies that interrogate parts of the cancer genome using bulk tumor samples consisting of millions of cells processed together. This strategy yields information regarding a tumor's global biology, but treatment efforts exploiting this technology have met with mixed results, suggesting the possibility of mixed cell populations within a single patient. Single cell analysis has confirmed this hypothesis in several cancers including glioblastoma and breast cancer [1].

Most high grade serous ovarian cancer (HGSOC) patients initially respond to platinum-based therapy, but the majority relapse and die from drug-resistant disease. It remains unclear if resistant clones are present early in tumor development or arise later, as a result of genomic instability or therapy-related genome damage. We hypothesized that ovarian cancer, which has both a high rate of chemotherapy-induced remission and a high rate of relapse, accompanied by increasing chemo-resistance, is likely composed of subsets of tumor cells with different gene expression patterns, biology, and chemosensitivity. Identification of subgroups with deleterious characteristics, such as stem cell or chemo-resistant signatures, could provide information on the pathways activated in these subgroups. This information could be clinically useful if there are available therapeutics targeting those pathways.

One characteristic that separates ovarian cancer from other common epithelial cancers is that ovarian cancers have very few recurrent mutations (eg. *TP53*) [2] . Instead, HGSOC is characterized by high levels of copy number alterations, which likely creates the genetic milieu responsible for the clinical aggressiveness and tendency to become resistant to

chemotherapy. Due to the lack of consistent driver mutations, many groups have used RNA sequencing and microarrays to identify activated pathways, classifying patients based on gene expression patterns. This work has identified four consensus HGSOC molecular subtypes, termed differentiated, proliferative, mesenchymal and immunoreactive [2]. Furthermore, correlation between molecular subtypes and response to treatment has been demonstrated [3].

A limitation of this analysis is that bulk sequencing of tumor samples is unable to identify small subpopulations of tumor cells, including cancer stem cells. In addition, sequences contributed by other cell populations in the bulk sample, including immune and stromal cells, can significantly alter gene expression patterns. For example, infiltrating immune and stromal cell subpopulations as low as 5% of the entire sample can bias molecular subtyping analyses [4, 5].

Single cell sequencing presents an alternative to bulk sequencing and may prove more useful in analyzing DNA and RNA alterations to define subpopulations and molecular targets of cancer cells for existing or novel therapeutics. In this study, we analyzed the transcriptome of 66 cells isolated from a single tumor specimen obtained during primary cytoreductive surgery.

2 Materials and methods

Detailed methods are available in Supplemental Methods.

2.1 Tissue and single cell preparation

Following approval from the University of Minnesota Institutional Review Board, tumor was collected from a patient with HGSOC during primary debulking surgery. The sample was dissociated into a single cell suspension and red blood cells were lysed. Remaining cells were fluorescently-labeled with an antibody cocktail targeting five immune cell markers: CD3, CD14, CD19, CD20, CD56 and sorted by flow cytometry. The non-immune population was sorted into individual wells using Fluidigm C1 chips and images of each cell were captured.

2.2 Sequencing and sequence processing

Isolation of mRNA and generation of barcode-labeled cDNA was performed on the Fluidigm C1 chip followed by sequencing using Illumina HiSeq2500. Sequences were mapped to the genome and transcriptome using Bowtie2 [6] and RNA-Seq by Expectation-Maximization [7]. Transcript analysis was limited to 24,200 RefSeq genes.

2.3 Sequence analysis

Unsupervised hierarchical and K-means clustering was done using Cluster 3.0 [8] and visualized using Treeview (v1.1.6r4) [9]. Principle component analysis was performed using R. Geneset enrichment analysis was performed using MSigDB v5.0 hallmark genesets [10, 11]. Comparisons to TCGA molecular subtypes was performed using custom R functions.

2.4 Immunohistochemistry

Standard methods were used for staining de-paraffinized and rehydrated tissue sections from the patient sample. Antibodies included ALDH1 (clone 44/ALDH), anti-PAX8 (clone MRQ-10), anti-p53 (clone DO7), anti-CD44(clone SP37), anti-CD133/Promin1(clone HA10), anti-Ki67 (clone SP6), and polyclonal cKIT(CD117).

3 Results

3.1 Cancer epithelial cells and cancer associated stromal cells can be differentiated based on gene expression patterns

Relative RNA abundance was calculated for 24,000 genes in 66 evaluable cells (Supp Table 1 and Supplemental Methods). We used three different methods of clustering cells based on their gene expression pattern and found that the 66 cells consistently separated into two groups (Fig 1). The three methods we used were unsupervised hierarchical clustering, K-means clustering, and principle components analysis.

To cluster cells, we selected the set of genes that had the highest variable expression based on an average absolute deviation > 3, encompassing 412 of the 4,673 highly expressed genes (Supp Tables 2 & 3). Unsupervised hierarchical clustering separated the cells into two major subsets as well as minor subsets (Fig 1A). K-means clustering (k = 4 for cells, k = 3 for genes) using the same set of genes separated the cells into the same two major groups, except for a single cell (Fig 1B). Principle component analysis using all 4,673 robustly expressed genes (Supp Table 2) also separated the cells into two major groups that correspond to the groups produced by hierarchical and K-means clustering (Figs. 1C & 1D). Importantly, bulk sequencing of the tumor would not have identified these distinct subsets of cells because the RNA from the low expressing and high expressing cells would be combined when processing the bulk sample.

We hypothesized that the two major groups resulted from separating cancer epithelial cells from cancer associated stromal cells. An alternative hypothesis, however, is that cells are clustering by cell cycle phase. To test the latter hypothesis, we performed the same clustering algorithms described above with a set of cell cycle genes defined by Whitfield, et al. [12] (Supp Table 3). Clustering exclusively using only cell cycle genes, or conversely, clustering using the remainder of the genes produced almost identical clusters to the original groupings, suggesting the cell cycle status of the cells does not explain the two groups (Supp Fig 1). We also attempted to define the cell cycle status of each cell using the mean-centered expression levels of cell cycle genes (Supp Table 4). By color-coding the cells based on cell cycle phase, it is evident that the clusters are not based on the cell cycle status of the cells (Supp Fig 2).

In support of the hypothesis that the two major groups are epithelial vs stromal cells we noted that 8 of 10 of the most differentially expressed genes are extracellular matrix (ECM) associated genes (Table 1). Approximately 76% of cells in the group highlighted in red in Figure 1 (16 of 21 cells) expressed high levels of ECM genes compared to the other group (Fig 2A). This same group of cells also express higher levels of the well-known mesenchymal transcription factors TWIST, SNAIL and ZEB (Fig 2B). It is possible that a

subset of cells in either of these groups are epithelial cancer cells undergoing an epithelial to mesenchymal transition (EMT), as several of these are EMT-associated proteins including collagen, fibronectin, secreted protein acidic and rich in cysteine, and vimentin [13]. In contrast to the red group with high ECM/EMT gene expression, 96% of the other group (43 of 45 cells), highlighted in blue in Figure 1, expressed PAX8 and/or CA125 (Fig 2C & 2D), which are markers indicative of epithelial ovarian cancer.

We performed gene set enrichment analysis (GSEA) to identify genesets that were enriched in either of the groups. The group of cells with high levels of ECM gene expression (group 2 in red in Fig 1) were significantly enriched for 8 out of 50 "hallmark" genesets annotated by the Broad Institute [14], while the other group (group 1 in blue in Fig 1) was enriched for three "hallmark" genesets (Table 2, Supp Fig. 3), (normalized enrichment score > 1.5 and nominal p value < 0.05). Not surprisingly, the top geneset for group 2, which is characterized by high ECM expression, was the EMT geneset in parallel with the overexpression of genes belonging to TGFb signaling and myogenesis genesets. The overexpression of these three genesets indicates the involvement of major cancer cell plasticity mechanisms characterizing this particular tumor. Also enriched were genesets for angiogenesis and hypoxia. While the other group of cells was significantly enriched for oxidative phosphorylation, DNA repair and MYC targets.

Two groups have published gene lists generated by comparing stromal tissue to epithelial tissue using either laser dissection on ovarian cancer samples [15] or by combining data from breast, ovarian and colorectal cancer [16]. Performing GSEA using these genesets indicates that the second group of cells, as expected, is highly enriched for genes that are “up in stroma”, while the first group of cells are highly enriched for genes that are “up in tumor” (Supp Fig 4). Again, it should be noted that these signatures would not have been identified using bulk sequencing data.

Analysis of single nucleotide variants (SNVs) detected by RNA sequencing also supports the hypothesis that group 1 cells are cancer epithelial cells while group 2 cells are stromal cells. Unfortunately, the RNA sequencing data was not deep enough to identify known recurrent mutations in ovarian cancer, like *TP53*. Instead, we used the RNASeq data to quantify numbers of homozygous vs heterozygous SNVs in each cell. After filtering out common SNVs, we identified ~650,000 SNVs that were present in at least one cell. The vast majority of these were unique to one or a few cells and were not “called” in most of the cells due to lack of sequencing reads covering the locus. Based on quantification of SNVs called as homozygous, cells could be grouped into low (< 15,000) and high (> 15,000) numbers of homozygous SNVs (Supp Figure 5A). As expected, the cells with low numbers of homozygous SNVs were mainly found in group 2, which we hypothesize are the cancer associated stromal cells (Supp Figure 5B), while the cells with high numbers of homozygous SNVs are in group 1, which we hypothesize are epithelial cancer cells. Henceforth, we will refer to these two groups as the stromal cell group (group 2 in red in Fig. 1) and epithelial cell group (group 1 in blue in Fig. 1). Interestingly, there is a subset of cells that have a high homozygous SNP count, yet cluster in the stromal group. It is possible that these cells represent cancer epithelial cells undergoing EMT.

3.2 IHC reveals the presence of stem cell markers and mesenchymal markers in both cancer and cancer associated stroma

To characterize protein expression in the ovarian cancer tumor sample that was used for single cell sequencing, we performed hematoxylin and eosin (H&E) staining and immunohistochemistry (IHC) for common ovarian cancer markers on paraffin embedded tumor sections that were adjacent to the piece used for single cell sequencing. Based on H&E staining, approximately half of the cells were tumor cells (outlined in Supp Figure 6A) and, as expected, these cells stained positively for the ovarian cancer cell markers PAX8 and TP53 (Supp Figure 6B & 6C). Approximately 20% of these tumor cells stained positively for the proliferation marker ki67 (MKI67) (Supp Figure 6D & 6E).

To evaluate the presence of stem cell markers, we performed IHC for CD117/cKIT, CD44, ALDH1A1, and CD133/PROM1[17–19]. There were a few cells with positive membrane staining for CD117/cKIT, mostly located within the stromal areas (Supp Figure 7A & 7B). A similar pattern was evident with CD44 staining (Supp Figure 7C & 7D), with 10% of stromal cells showing immunoreactivity, while only 1% of tumor cells stained positively (Supp Figure 7E). ALDH1A1 staining was pervasive throughout the tumor and stromal areas, with cytoplasmic staining evident in a subset of tumor cells, while very few cells stained positively for CD133/PROM1 (Supp Figure 8). The vast majority of stromal cells were positive for vimentin (VIM), while only a few cells within the tumor sections stained positively (Supp Figure 9).

3.3 Technical considerations for single cell sequencing

We were able to sequence cDNA generated from 92 single cells using 100 bp paired-end sequencing to an average depth of 1.8 million reads/cell. An average of 89% (range 73% to 96%) mapped to the genome, but only 26% (range 1% to 47%) aligned to the transcriptome, compared to 46% for the bulk sample (Supp Table 5). This might indicate that ovarian cancer cells produce a large amount of mRNA that is not processed correctly or the process of generating cDNA from single cells captures some genomic DNA. The number of RefSeq genes with detectable reads in the bulk sample was 12,193, while the average for a single cell was 3,603 (range 1,735 – 5,929). This suggests that either each single cell expresses about 25% of the genes compared to the total number of genes expressed in the bulk sample, or the technical limitations of single cell RNA capture and processing results in a large number of expressed genes not being detected.

An analysis of housekeeping genes, which we hypothesized to be robustly expressed and detected in all cells, indicated that some housekeeping genes were expressed in all cells (e.g. *ACTB* and *CTNNB1*), while other prototypical housekeeping genes (e.g. *GAPDH* and *TBP*) were not detected in all cells. In general, however, the entire set of 4,673 robustly expressed genes was highly enriched for housekeeping genes (1,924 out of 4,673, Fisher's Exact Test, $P < 0.00001$), based on a list of 3,873 housekeeping genes identified by Eisenberg, et al., [20] (Supp Table 3). This suggests that technical limitations can partially explain why low numbers of genes are detected in single cells. Nevertheless, it is apparent that there is extensive heterogeneity in gene expression when comparing single cells, as the majority of

the 4,673 highly expressed genes are detected in less than half of the 66 cells sequenced (Supp Fig 10).

3.4 Detecting a chemo resistant population

To determine whether or not a chemo resistant gene expression signature is present in either the stromal or epithelial group, we performed GSEA using genesets produced by three different studies that identified genesets that were upregulated or downregulated in platinum resistant ovarian cancer (Supp Table 6). The first geneset was derived from cell lines conditioned to be resistant to either cisplatin or paclitaxel [21]. Both the cisplatin and paclitaxel resistant downregulated genesets were upregulated in the stromal group (FDR < 25%), while the corresponding resistant upregulated genesets were enriched in the epithelial group, although not to a significant level (Supp Fig 11A & B). We had expected to find enrichment for genes upregulated in platinum resistant cancers in the stromal group, based on the finding that patients classified in the “mesenchymal” molecular subtype have a worse prognosis than the other groups [3] and the general consensus that EMT cells are more resistant to chemotherapy [22]. Next, we performed GSEA using genesets that corresponded to genes upregulated or downregulated in platinum-sensitive vs platinum-resistant tumor samples. The two genesets were derived from the union of genes found upregulated or downregulated in resistant vs sensitive ovarian cancer patients and genes that were changed in an ovarian cancer cell line (IGROV1) compared to its carboplatin-resistant derivative [23]. The only significant finding was an enrichment in the epithelial group for genes that were downregulated in carboplatin resistance tumors (Supp Fig 11C). Finally, we used a list of genes produced by the International Cancer Genome Consortium (ICGC) that were significantly upregulated or downregulated when comparing resistant vs sensitive ovarian cancer patient samples [24]. The only significant association was between the epithelial group and genes upregulated in platinum resistant cancers (Supp Fig 11D). One caveat to these analyses are that there is minimal overlap between the genesets defined in the three different studies as significantly up or downregulated in platinum resistant cells/tissues. Based on these GSEA results we conclude that neither group displays a gene signature that corresponds to platinum and taxane resistance based on the studies described above.

3.5 Detecting cancer stem cells

We analyzed the expression of cancer stem cell markers in our dataset (Fig 3). The patient in this study had stage IIIC HGSOC and approximately 30% (19/66) of her single cells were CD44+, although the majority (15) of these cells were in the stromal group (Fig 3A). This level of CD44 positivity is correlated with poor survival in early stage ovarian patients, but does not have a correlation with survival in late stage ovarian cancer patients [25]. There were no individual cells that were double positive for CD44 and KIT, nor were there any cells double positive for PROM1 and ALDH1, two combinations that have been purported to define ovarian cancer stem cells [17, 19, 26] (Fig 3A, D, E & H). In general, the stromal group was enriched for CD44, while the epithelial group was strongly enriched for CD24 (Fig 3A & B). Interestingly, EPCAM, a gene associated with platinum resistance and ovarian cancer metastases [27] was unexpectedly correlated with CD44-/CD24+ cells (Fig 3C), although four cells were triple positive for CD44/CD24/EPCAM (Fig 3A–C). MYD88 and CD44 expression were not correlated, and high expression of CK18 was negatively

correlated with CD44 expression (Fig 3A, G & I). Only a few cells expressed PROM1 and KIT and there was no overlap between ALDH1+, PROM1+ or KIT+ cells (Fig 3D–F). Interestingly, there was a single cell that simultaneously expressed CD44, MYD88, ALDH1, and high levels of CK18, representing a candidate stem cell based on co-expression of these markers [28, 29] (black arrows, Fig 3A, F, G & I).

3.6 Molecular classification of ovarian cancer comparing bulk to single cell gene expression patterns

We re-derived the TCGA patient molecular subtypes using K-means clustering using TCGA RNASeq data instead of using the microarray data that was used originally in the TCGA study (see Materials and Methods). Surprisingly, approximately 30% of patients would be assigned to different molecular subtypes based on RNASeq data (Supp Fig 12), including almost half of the immunoreactive group. To determine the molecular subtype of the single cells, bulk population, and the stromal and epithelial cell groups we included these as samples along with the TCGA patient RNASeq data and performed K-means clustering. Using the RNASeq data from the bulk population, the patient's bulk sample is classified as a mesenchymal subtype. As might be expected, the stromal group of cells are also classified as mesenchymal, while the epithelial group are classified as proliferative. Analysis of each single cell classifies the majority of stromal cells as mesenchymal and epithelial cells as proliferative, although some of the cells are classified in other groups (Fig 4). These analyses indicate that molecular classification using bulk sequencing can mask the gene expression patterns of large groups of cells, which may ultimately effect patient outcome and possible treatment decisions based on these classifications.

3.7 Functional classification of single cells

We explored the possibility of functionally classifying single cells based on genes that have been defined in the literature as marker genes for stroma and stromal subgroups and for epithelial cancer cells and epithelial subgroups. We used these known functional markers to define cancer epithelial cells vs stroma cells and to create three subgroups of cancer (epithelial, EMT, and EMP) and four subgroups of stroma (fibroblasts or myofibroblasts, either activated or non-activated). A cell was defined as a "cancer" cell (n = 54) if it expressed one or more ovarian cancer marker gene (*WT1*, *PAX8*, *MUC16*, *KL6*, *KL7* or *KL8*). The "cancer" group was further subdivided based on epithelial markers (*CD24*, *EPCAM*, *CDH1*, *OCLN*, *KRT19*, *DSP*, *KRT18*, *CLDN4*) or mesenchymal markers (*CD44*, *CDH2*, *ITGA5*, *VIM*, *FN1*, *S100A4*, *TNC*, *MMP2*, *ACTA2*, *TWIST1*, *WNT5A*, *SNAI2*, *ZEB1*, *ZEB2*). Cells strongly expressing at least two genes from a marker group were considered to present an epithelial or EMT signature. Many cells expressed both signatures and were classified as being in an epithelial-mesenchymal plasticity state (EMP) [30]. Four of the 12 cells that were negative for cancer markers also expressed both the epithelial and mesenchymal signatures and were classified as cancer-negative EMP cells. The remaining eight cells were defined as fibroblasts or myofibroblasts based on *ACTA2* levels and as activated or non-activated based on *FAP* expression. By overlaying these groups on the PCA plot it is evident that fibroblasts cluster in the stromal group while the EMP/EMT/epithelial cells cluster in the epithelial group (Fig. 5). Interestingly, the single cell displaying the most stem cell markers in Fig. 3 is classified as a non-cancer EMP cell in this grouping.

Discussion

In this study of HGSOC we identified two major groups of cells, which were characterized by stromal and epithelial gene expression signatures. Neither of these groups displayed gene expression patterns associated with chemo resistance based on three independent studies [21, 23, 24]. However, the chemo resistant genesets produced by these three studies did not overlap, indicating they may not be true indicators of chemo resistance. The patient in this study has shown no evidence of recurrence 19 months post-surgery which is consistent with the finding that the single cells did not express a chemo-resistant gene signature. Analysis of single cells from more patients, including samples from patients before and after recurrence will be necessary to define chemo-resistant single cell signatures. This type of analysis will also help answer the question of whether or not the resistant cell type was present in the primary tumor.

Identifying the ovarian cancer stem cell will likely be crucial for improving current cure rates of less than 50% for advanced stage patients. Many studies have attempted to identify ovarian cancer stem cells, however, molecular markers that indisputably identify ovarian cancer stem cells are not well-defined [31–33]. The consensus is that the cancer stem cell population is rare (< 2%) [31, 32], although this might be an underestimate due to the technical difficulty of propagating cancer stem cells [34]. Future studies will be necessary to quantify the frequency of cells with stem cell markers in other HGSOCs and sorting these cells followed by functional analyses will be required to determine their stemness.

Clinical decision-making based on molecular subtyping using gene expression patterns is still a rarity in oncology, except in a few types of cancers, like breast cancer. One reason may be that the cell types responsible for chemo resistance and/or recurrence are rare and their gene signature is always masked when analyzing gene expression data from a bulk tumor sample. Often, the molecular subtypes defined by gene expression patterns do not correlate with survival or have predictive value for alternative treatment options. In ovarian cancer, TCGA and other groups used clustering algorithms to define four molecular subtypes, referred to as mesenchymal, immunoreactive, proliferative and differentiated based on key genes that are expressed in each subtype. These uniquely defined molecular subtypes have some prognostic relevance and possible differential response to antiangiogenic treatment with bevacizumab [2, 3, 15, 35]. However, when the same clustering analysis is performed using bulk RNASeq data, which was gathered after the initial TCGA ovarian cancer publication, approximately 30% of patients are classified in different groups than they were originally classified when using microarray data (Supp Fig 12). Based solely on the bulk sample, the patient in this study falls into the mesenchymal subtype. However, if we classify the stromal group of cells and the epithelial group of cells as two “patients”, the stromal group is classified as mesenchymal while the epithelial group is classified as proliferative. Finally, if we classify each single cell, the majority of stromal cells fall into the mesenchymal group and the majority of the epithelial cells fall into the proliferative group, although a small portion of each group are classified in one of the other three groups (Fig 4). Interestingly, this patient’s cancer harbors the molecular subtypes, mesenchymal and proliferative, which carry the worst prognosis, but which may derive the largest clinical benefit from anti-VEGF treatment with bevacizumab. These results suggest

that molecular classification at the single cell level can reveal the presence of multiple molecular subclassifications within a tumor. This knowledge may prove useful for categorizing subpopulations of cancer cells which drive clinical outcome and response to chemotherapy, although more patients will need to be analyzed.

In summary, to our knowledge, we provide the first study of gene expression patterns in single cells from a patient with HGSOC, demonstrating there is considerable heterogeneity within a single tumor. Analysis of more patients will allow stratification based on percentages of cells in the different functional groups. There appear to be underlying themes within this heterogeneity that can be defined at the single cell level and it is tantalizing to hypothesize that it may be possible to identify cancer stem cells using this technology. Our findings provide a first view of single cell gene expression analysis in ovarian cancer. Analysis of more patient samples and more cells per sample will be required to continue to unravel the complexity of HGSOC and aid in development of effective therapies tailored to individual patients with the goal of improving outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grant Support

This work was supported by the following funding sources: Masonic Cancer Center, University of Minnesota (TWG grant to TKS, MB, and MAB); NCI/NIH Project grant (P30-CA77598); The National Center for Advancing Translational Sciences of the NIH (UL1TR000114). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

The authors would like to acknowledge the contributions of Colleen Forster from the University of Minnesota Histology and IHC Laboratory, the University of Minnesota Genomics Center for Fluidigm reagents, Nisha Shah from the University of Minnesota Flow Core, and the Minnesota Supercomputing Institute. We would like to thank Dr. Alessia Zoso from the University of Miami for providing her expertise on separation of immune cells.

References

1. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science. 2014; 344:1396–1401. [PubMed: 24925914]
2. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. Nature. 2011; 474:609–615. [PubMed: 21720365]
3. Konecny GE, Wang C, Hamidi H, Winterhoff B, Kalli KR, Dering J, et al. Prognostic and therapeutic relevance of molecular subtypes in high-grade serous ovarian cancer. J Natl Cancer Inst. 2014:106.
4. Calon A, Lonardo E, Berenguer-Llergo A, Espinet E, Hernando-Momblona X, Iglesias M, et al. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. Nat Genet. 2015; 47:320–329. [PubMed: 25706628]
5. Isella C, Terrasi A, Bellomo SE, Petti C, Galatola G, Muratore A, et al. Stromal contribution to the colorectal cancer transcriptome. Nat Genet. 2015; 47:312–319. [PubMed: 25706627]
6. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology. 2009; 10:R25. [PubMed: 19261174]
7. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12:323. [PubMed: 21816040]

8. de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. Bioinformatics. 2004; 20:1453–1454. [PubMed: 14871861]

9. Page RD. TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci. 1996; 12:357–358. [PubMed: 8902363]

10. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102:15545–15550. [PubMed: 16199517]

11. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003; 34:267–273. [PubMed: 12808457]

12. Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, et al. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol Biol Cell. 2002; 13:1977–2000. [PubMed: 12058064]

13. Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial-mesenchymal transition. Nat Rev Mol Cell Biol. 2014; 15:178–196. [PubMed: 24556840]

14. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011; 27:1739–1740. [PubMed: 21546393]

15. Tothill RW, Tinker AV, George J, Brown R, Fox SB, Lade S, et al. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. Clinical cancer research : an official journal of the American Association for Cancer Research. 2008; 14:5198–5208. [PubMed: 18698038]

16. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nature communications. 2013; 4:2612.

17. Zhang S, Balch C, Chan MW, Lai HC, Matei D, Schilder JM, et al. Identification and characterization of ovarian cancer-initiating cells from primary human tumors. Cancer Res. 2008; 68:4311–4320. [PubMed: 18519691]

18. Curley MD, Therrien VA, Cummings CL, Sergent PA, Koulouris CR, Friel AM, et al. CD133 expression defines a tumor initiating cell population in primary human ovarian cancer. Stem Cells. 2009; 27:2875–2883. [PubMed: 19816957]

19. Silva IA, Bai S, McLean K, Yang K, Griffith K, Thomas D, et al. Aldehyde dehydrogenase in combination with CD133 defines angiogenic ovarian cancer stem cells that portend poor patient survival. Cancer Res. 2011; 71:3991–4001. [PubMed: 21498635]

20. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. Trends Genet. 2013; 29:569–574. [PubMed: 23810203]

21. Sherman-Baust CA, Becker KG, Wood WH Iii, Zhang Y, Morin PJ. Gene expression and pathway analysis of ovarian cancer cells selected for resistance to cisplatin, paclitaxel, or doxorubicin. J Ovarian Res. 2011; 4:21. [PubMed: 22141344]

22. Polyak K, Weinberg RA. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. Nat Rev Cancer. 2009; 9:265–273. [PubMed: 19262571]

23. Cheng L, Lu W, Kulkarni B, Pejovic T, Yan X, Chiang JH, et al. Analysis of chemotherapy response programs in ovarian cancers by the next-generation sequencing technologies. Gynecol Oncol. 2010; 117:159–169. [PubMed: 20181382]

24. Patch AM, Christie EL, Etemadmoghadam D, Garsed DW, George J, Fereday S, et al. Whole-genome characterization of chemoresistant ovarian cancer. Nature. 2015; 521:489–494. [PubMed: 26017449]

25. Steffensen KD, Alvero AB, Yang Y, Waldstrom M, Hui P, Holmberg JC, et al. Prevalence of epithelial ovarian cancer stem cells correlates with recurrence in early-stage ovarian cancer. J Oncol. 2011; 2011:620523. [PubMed: 21904548]

26. Kryczek I, Liu S, Roh M, Vatan L, Szeliga W, Wei S, et al. Expression of aldehyde dehydrogenase and CD133 defines ovarian cancer stem cells. Int J Cancer. 2012; 130:29–39. [PubMed: 21480217]

27. Bellone S, Siegel ER, Cocco E, Cargnelutti M, Silasi DA, Azodi M, et al. Overexpression of epithelial cell adhesion molecule in primary, metastatic, and recurrent/chemotherapy-resistant

epithelial ovarian cancer: implications for epithelial cell adhesion molecule-specific immunotherapy. Int J Gynecol Cancer. 2009; 19:860–866. [PubMed: 19574774]

28. Alvero AB, Chen R, Fu HH, Montagna M, Schwartz PE, Rutherford T, et al. Molecular phenotyping of human ovarian cancer stem cells unravels the mechanisms for repair and chemoresistance. Cell Cycle. 2009; 8:158–166. [PubMed: 19158483]
29. Landen CN Jr, Goodman B, Katre AA, Steg AD, Nick AM, Stone RL, et al. Targeting aldehyde dehydrogenase cancer stem cells in ovarian cancer. Mol Cancer Ther. 2010; 9:3186–3199. [PubMed: 20889728]
30. Ye X, Weinberg RA. Epithelial-Mesenchymal Plasticity: A Central Regulator of Cancer Progression. Trends Cell Biol. 2015; 25:675–686. [PubMed: 26437589]
31. Shah MM, Landen CN. Ovarian cancer stem cells: are they real and why are they important? Gynecol Oncol. 2014; 132:483–489. [PubMed: 24321398]
32. Burgos-Ojeda D, Rueda BR, Buckanovich RJ. Ovarian cancer stem cell markers: Prognostic and therapeutic implications. Cancer letters. 2012
33. Bast RC Jr, Hennessy B, Mills GB. The biology of ovarian cancer: new opportunities for translation. Nat Rev Cancer. 2009; 9:415–428. [PubMed: 19461667]
34. Quintana E, Shackleton M, Sabel MS, Fullen DR, Johnson TM, Morrison SJ. Efficient tumour formation by single human melanoma cells. Nature. 2008; 456:593–598. [PubMed: 19052619]
35. Verhaak RG, Tamayo P, Yang JY, Hubbard D, Zhang H, Creighton CJ, et al. Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. J Clin Invest. 2013; 123:517–525. [PubMed: 23257362]

Highlights

- A single tumor contains cells representing all defined molecular classifications
- Molecular classification based on sequencing of bulk tumor samples is problematic
- Single cell sequencing can identify rare cells expressing stem cell markers

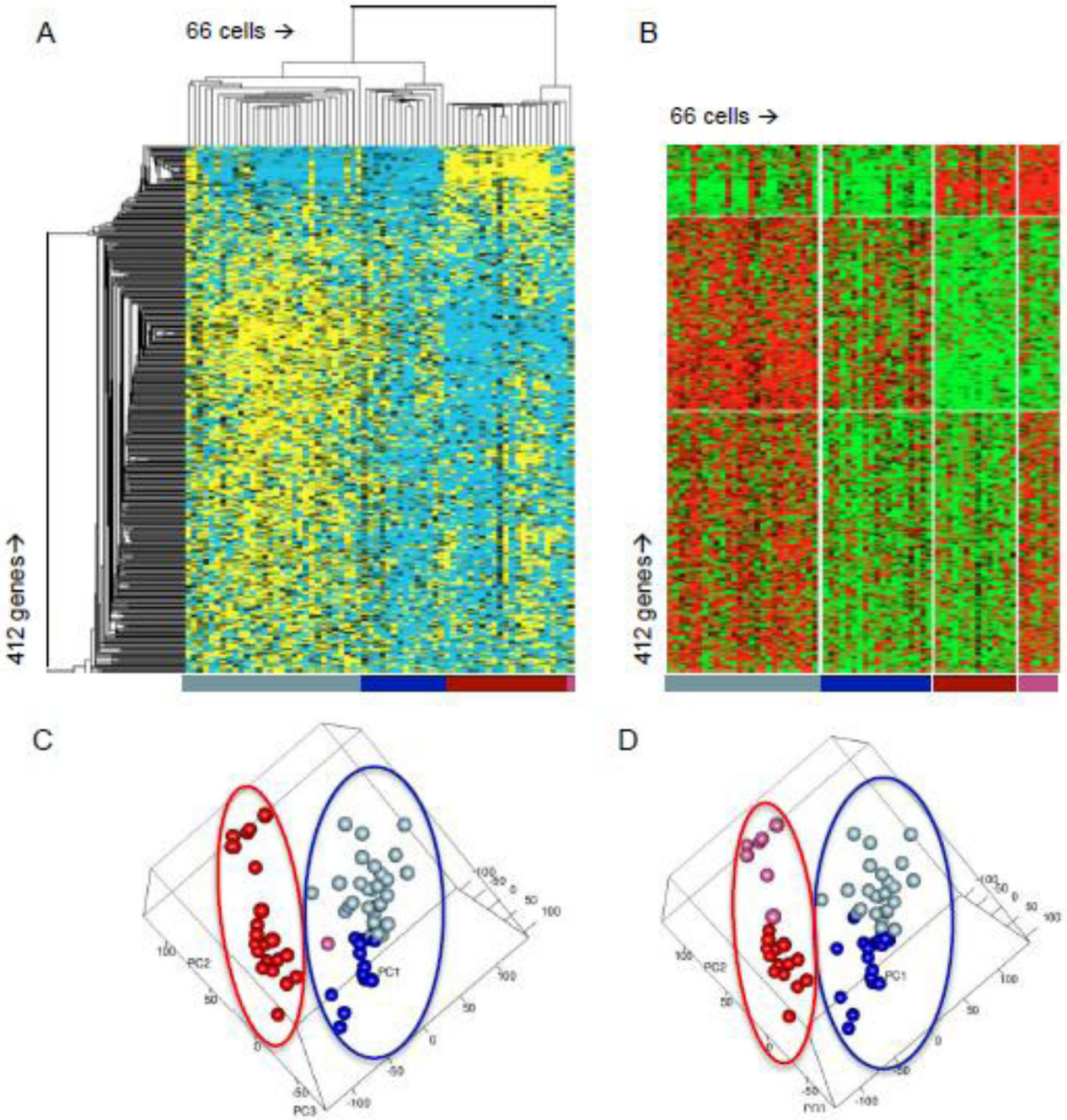

Figure 1.

Three clustering methods reveal two major groups of cells. A) Unsupervised hierarchical clustering based on 412 variably expressed genes. B) K-means clustering using the same 412 genes. C & D) Principle component analysis (PCA) 3D plot of 66 cells based on first three principle components analysis using 4,673 highly expressed genes. Cells are colored in C) based on color bar underneath hierarchical clustering heat map. Cells in D) are colored based on K-means color bar. The two major groups defined by all three methods (Group 1 = blue/lightblue vs Group 2 = red/lightred) are identical except for a single cell.

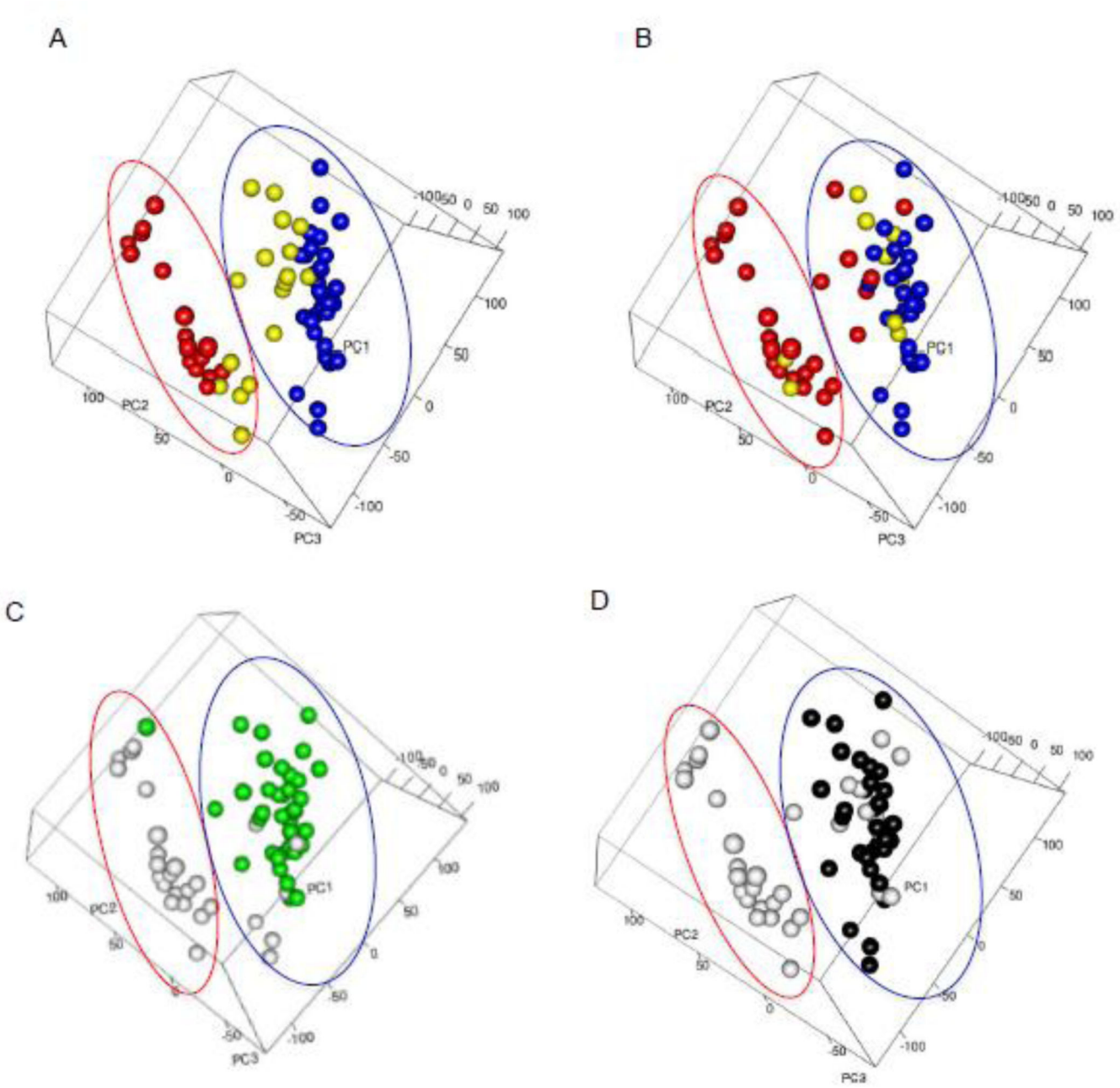

Figure 2.

High expression levels of 8 ECM-related-genes define cell subpopulations. PCA plots using 4,673 genes × 66 samples with cells colored based on average expression levels of A) 8 highly variably expressed ECM genes, and B) EMT transcription factors (TWIST, SNAIL, ZEB). Red = high, yellow = medium, blue = low. C) PAX8 high in green. D) CA125 (MUC16) high in black.

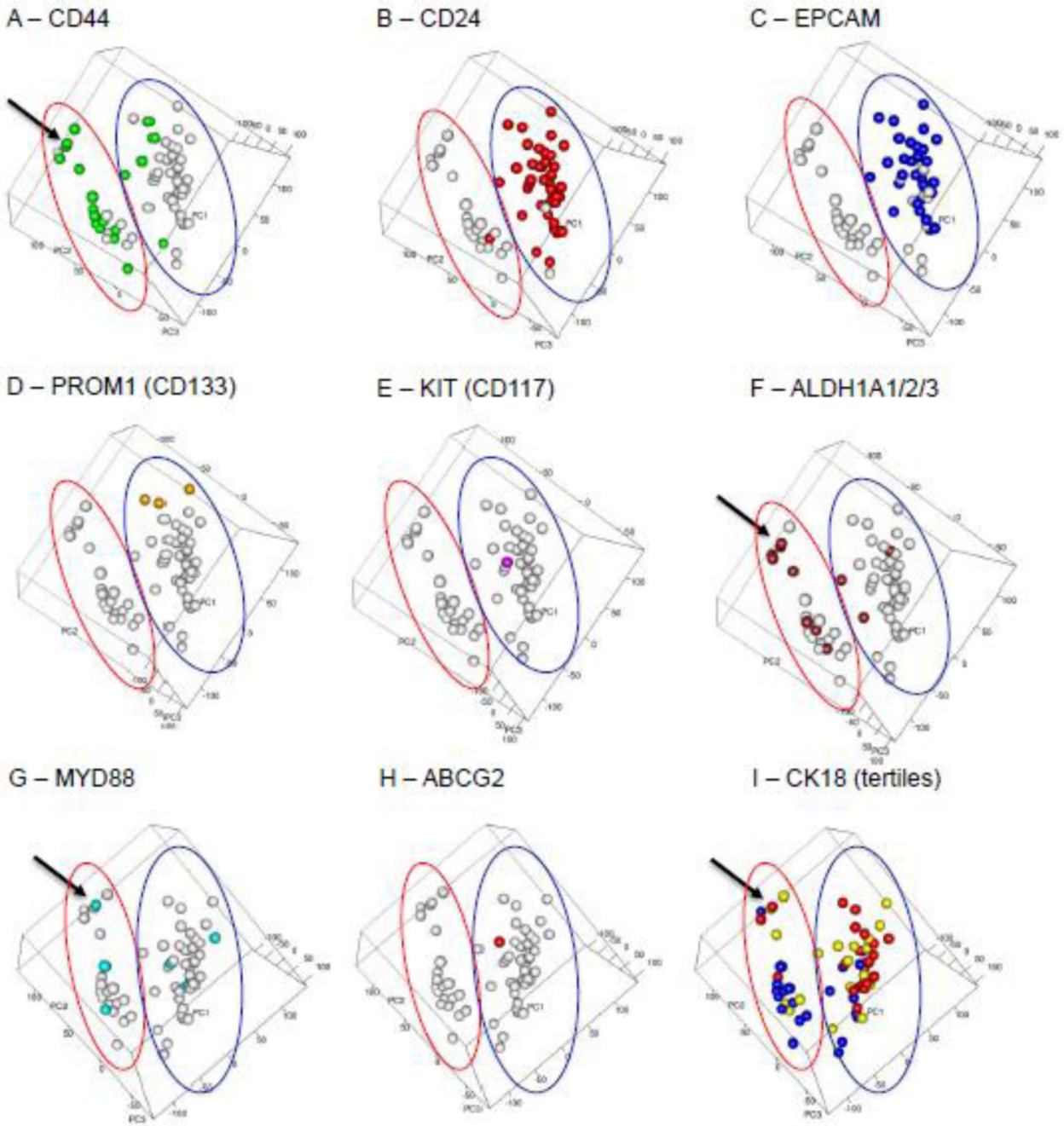

Figure 3.

PCA plots of all 66 cells colored based on expression of the indicated stem cell gene. A) CD44 in green. B) CD24 in red. C) EPCAM in blue. D) PROM1 in yellow. E) KIT in pink. F) ALDH1A1, ALDH1A2 and ALDH1A3 in brown. G) MYD88 in cyan. H) ABCG2 in red. In A-H, cells are colored if expression of the indicated gene(s) is > 2 (LogTPM), otherwise cells are in white. I) CK18 expression based on expression tertiles: low/none = blue, middle = yellow, high = red. Red and blue circles delineate stroma and epithelial subgroups identified in Figure 1, respectively.

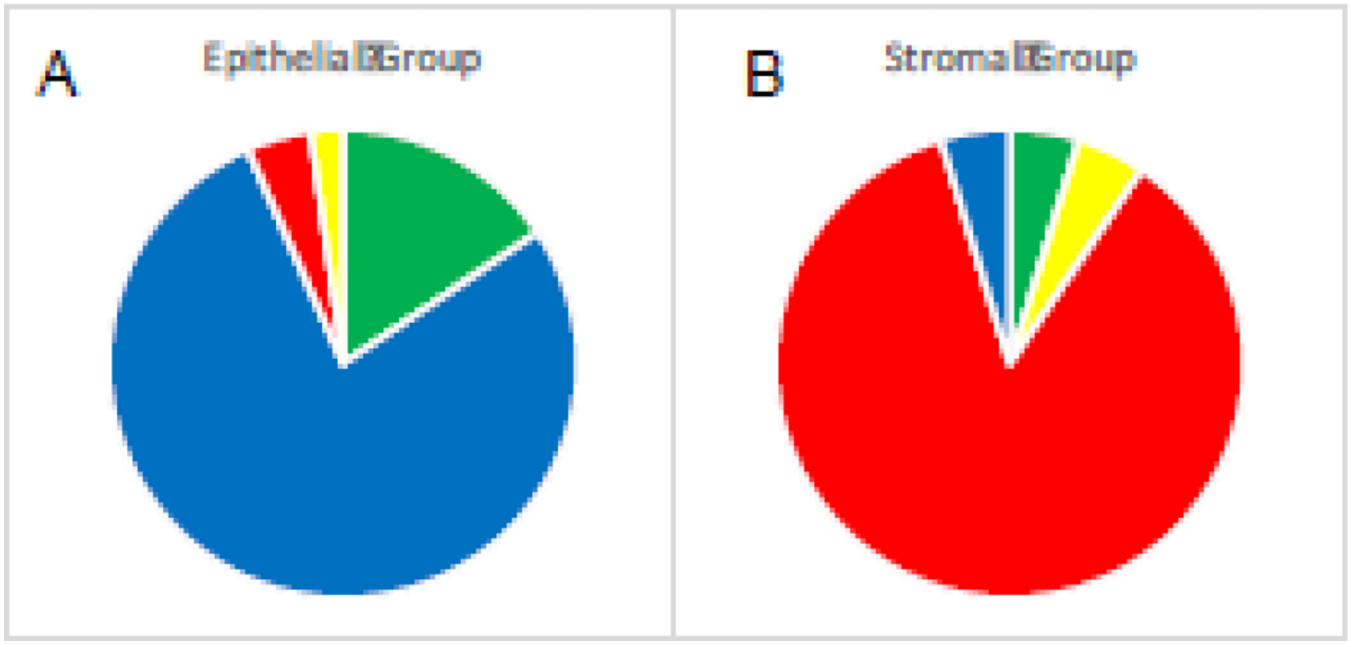

Figure 4.

Pie charts depicting distribution of single cells based on molecular subtype. A) Epithelial group cells (n=45) and B) Stroma group cells (n=21). Differentiated = green, proliferative = blue, mesenchymal = red, immunoreactive = yellow.

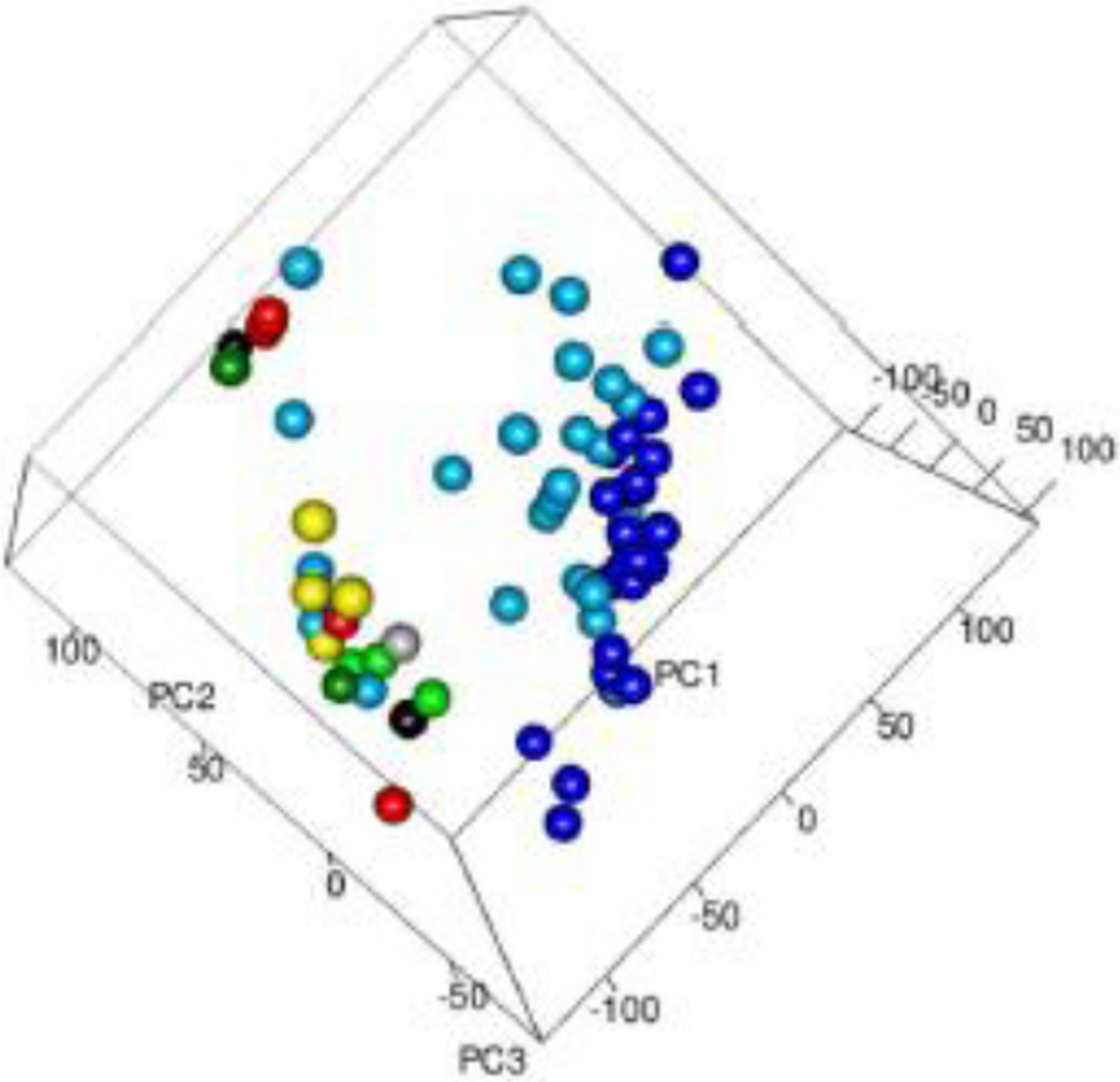

Figure 5.

PCA plot with single cells colored based on presence of functional markers: Cancer epithelial cells (dark blue), cancer EMP cells (blue), cancer EMT cells (yellow), non-cancer EMP cells (red), fibroblasts (activated = black, not activated = grey), and myofibroblasts (activated = dark green, not activated = light green).

Table 1

Top ten variably expressed genes*

HGNC_symbol	Name	Function
C0L3A1	Collagen, Type III, Alpha 1	ECM protein
SPARC	Secreted Protein, Acidic, Cysteine-Rich (Osteonectin)	Matrix-associated protein that regulates collagen, cell shape, and ECM synthesis
COL1A1	Collagen, Type 1, Alpha 1	ECM protein
COL1A2	Collagen, Type 1, Alpha 2	ECM protein
DCN	Decorin	Proteoglycan that plays a role in matrix assembly
VIM	Vimentin	Intermediate filament that maintains cell shape and stabilizes type 1 collagen
CD24	CD24 Molecule	Membrane sialoglycoprotein involved in cell signaling
FN1	Fibronectin 1	Glycoprotein involved in cell adhesion, motility and matrix assembly
ARRDC3	Arrestin Domain Containing 3	Membrane protein implicated in B2- adrenergic receptor downregulation
THBS1	Thrombospondin 1	Adhesive glycoprotein that mediates cell-cell and cell-matrix adhesion

*Variation based on average of absolute deviation of mean-centered log2TPM values

Table 2

Gene set enrichment comparing Group 1 (blue) cells to Group 2 (red) cells

Geneset	Enriched Group	SIZE[1]	NES[2]	NOM p-val[3]	FDR q-val[4]	FWER p-val[5]
OXIDATIVE PHOSPHORYLATION	Group 1 (blue)	142	1.80	0.01	0.06	0.05
DNA REPAIR	Group 1 (blue)	59	1.62	0.01	0.18	0.24
MYC TARGETS VI	Group 1 (blue)	165	1.56	0.04	0.20	0.34
EPITHELIAL MESENCHYMAL TRANSITION	Group 2 (Red)	122	−2.26	0.00	0.00	0.00
UV RESPONSE DOWN	Group 2 (Red)	67	−2.08	0.00	0.00	0.00
ANGIOGENESIS	Group 2 (Red)	16	−2.03	0.00	0.00	0.00
MYOGENESIS	Group 2 (Red)	54	−1.95	0.00	0.00	0.01
HYPOXIA	Group 2 (Red)	101	−1.88	0.00	0.00	0.02
IL2 STATS SIGNALING	Group 2 (Red)	75	−1.86	0.00	0.00	0.03
COAGULATION	Group 2 (Red)	52	−1.70	0.01	0.02	0.15
TGF BETA SIGNALING	Group 2 (Red)	32	−1.61	0.03	0.04	0.25

[1] Number of genes in geneset after filtering genes not in expression dataset

[2] Normalized enrichment score is the enrichment score for the geneset after it has been normalized across analyzed genesets

[3] Nominal p value is the statistical significance of the enrichment score

[4] False discovery rate. Probability that the NES represents a false positive finding

[5] Familywise-error rate. A more conservative estimated probability that the NES represents a false positive finding