Gynecologic Oncology xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Gynecologic Oncology

journal homepage: www.elsevier.com/locate/ygyno

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

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.

ARTICLE INFO

Article history:
Received 1 December 2016
Received in revised form 9 January 2017
Accepted 11 January 2017
Available online xxxx

Keywords:
Ovarian cancer
Molecular subtypes
Single cell sequencing and cancer stem cells

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 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.

1. Introduction

The promise of individualized cancer therapy is predicated on the identification of remediable drug targets within a tumor. Target

* Corresponding author at: MMC 395, 420 Delaware St SE, Minneapolis, MN 55455, USA.
E-mail address: star0044@umn.edu (T.K. Starr).

http://dx.doi.org/10.1016/j.ygyno.2017.01.015
0090-8258/

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 (e.g. *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 *C*1 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 4673 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 4673 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 (Fig. 1C & D). 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 Fig. 1 (16 of 21 cells) expressed high levels of ECM genes compared

Please cite this article as: B.J. Winterhoff, et al., Single cell sequencing reveals heterogeneity within ovarian cancer epithelium and cancer associated stromal cells, Gynecol Oncol (2017), http://dx.doi.org/10.1016/j.ygyno.2017.01.015