Differentially expressed markers between chemoresistant and non-chemoresistant cells in TNBC

Xinhe Zheng
College of Letters & Science, UC Berkeley, 110 Sproul Hall Berkeley, CA 94720
xinhezheng@hotmail.com

Abstract. This study focuses on chemoresistance in triple-negative breast cancer (TNBC). Dataset GSE118389 and paper Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq from Simona Cristea from Dana-Farber Cancer Institute was used to discover genetic markers that are related to chemoresistance in TNBC. Differentially expressed genes between a chemoresistant cluster and non-chemoresistant clusters in various subpopulations of the TNBC samples were identified, compared, and analyzed. 1987 markers that are differentially expressed in the chemoresistant cluster and non-chemoresistant clusters were found.

Keywords: triple-negative breast cancer, chemoresistance, single-cell RNA sequencing, differential expression test.

1. Introduction
Breast cancer is a prevalent disease around the world. In 2020, there are approximately 2.26 million women newly diagnosed with breast cancer, and 684,996 deaths caused by breast cancer in women [1]. Early-stage, non-metastatic (or metastasis to lymph nodes only) breast cancer is curable in 70-80% of patients whereas advanced breast cancer with metastases is considered incurable. 10% of breast cancers, including those with BRCA1, BRCA2 mutations, are inherited, and about 20% are risk modifiable, depending on the level of obesity, physical activity, and alcohol use. There are 3 molecular types in breast cancer: activation of human epidermal growth factor receptor 2 (HER2), activation of hormone receptors (oestrogen and progesterone, e.g., ER), and BRCA mutations. There are 5 subtypes of breast cancers: Triple-negative, HER2-enriched (non-luminal), Luminal B-like HER2+, Luminal B-like HER2-, and Luminal A-like [2].

This paper focuses on triple-negative breast cancer (TNBC). TNBC includes cells without receptors for oestrogen, progesterone and HER2 protein. When these proteins attach to their respective receptors in cancer cells, cancer cells are triggered to grow. TNBC is uncommon: only approximately 15% of breast cancers are triple-negative [3].

Another focus of this paper is chemotherapy resistance. When cancers stop responding to treatment, they are said to be resistant to chemotherapy. [4]. Therapeutic resistance that occurs in cancer stem cells (CSCs) is a probable cause of metastatic relapse after chemotherapy since their avoidance of apoptosis allows the tumor to re-develop after the therapy [5]. Various intracellular mechanisms that enable cells...
to evade the cytotoxic effects of therapeutics are related to CSCs: for example, the regulation of drug availability, the epithelial–mesenchymal transition (EMT), and oncogenic signaling pathways [6].

This study is on chemoresistance in TNBC. Dataset GSE118389 is used to discover genetic markers that are related to chemoresistance in TNBC [7]. All data analysis was performed in R. The R package Seurat from the Satija lab was used to process and analyze the dataset to find differentially expressed genes between a chemoresistant cluster and non-chemoresistant clusters in various subpopulations of the TNBC samples. The purpose of this study is to identify chemoresistant markers and explore their significance.

2. Literature review

There are not many studies on chemoresistant markers in TNBC in the early years, but thanks to the development of genome sequencing, there are more and more studies nowadays. These studies mainly concentrate on the therapeutics or the mechanism of resistance of one or a few related genes.

Several papers explore the roles of genes involved in the chemoresistance of TNBC with a focus on therapeutics. TNFSF13 upregulation is linked to a poor chemotherapy response in TNBCs, which can be alleviated by therapeutic targeting of autophagy initiation [8]. Drug therapy causes apoptotic cell death in chemoresistant cells when DSTYK is knocked out using the CRISPR/Cas9 technology [9]. Chemotherapy increased the expression of PTN and PTPRZ1, which is regulated by CDKN1A. The NF-B pathway was involved in chemoresistance caused by chemotherapy-induced increases in the CDKN1A/PTN/PTPRZ1 axis [10]. In addition, STAT3 was identified as a marker associated with DOX resistance in the CD44+/high/CD24-/low/ALDH1+ BCSCs-like subpopulation [11]. CSC and chemoresistant TNBC cells are resensitized to chemotherapeutic drugs when RAD50 within the Mre11-Rad50-NBS1 (MRN) complex is depleted or blocked [12]. And, MLK4 regulates the pro-survival response to DNA-damaging therapies, which promotes TNBC chemoresistance [13]. Overexpression of ATP-binding cassette (ABC) transporter proteins, such as P-glycoprotein (Pgp) causes TNBCs to mediate drug resistance [14]. The elevated expression of AR and GATA3 characterize the TNBC subgroup that is predicted to have a relatively favorable prognosis [15]. Also, through the IL-6-TGF-1/IGT1 axis, the interaction between TNBC cells and tumor-associated macrophages (TAMs) promotes persistent activation of HLF in tumor cells, which speeds up the development of malignant tumors by encouraging ferroptosis resistance in TNBC cells [16].

Other papers discover the mechanism of various genes related to TNBC chemoresistance. The dysregulation of the baculoviral IAP repeat-containing family genes' expression levels in cancer tissue, which code for inhibitors of apoptosis proteins, suggests that TNBC's growth and chemoresistance to treatment may be due to a disrupted apoptosis process in cancer cells [17]. ADAM10 knockdown resulted in a significant reduction in cell proliferation, migration, invasion, and paclitaxel and adriamycin IC50 values, as well as cell cycle arrest and apoptosis, which were linked to Notch signaling, CD44, and cellular prion protein downregulation (PrPc) [18]. Chemoresistant cells had significantly higher FSTL1 levels. In breast cancer cells, a miR-137/FSTL1/integrin 3/Wnt/β-catenin signaling axis modulates chemoresistance and stemness [19]. Upregulation of p-Akt/p-GSK-3 (Ser9)/β-catenin/survivin after Glut1 ablation reduced apoptosis and elevated drug resistance [20]. Additionally, by up-regulating the Rac1/β-catenin pathway, TUFT1 promotes tumor cell metastasis and stemness. [21]. Chemoresistant TNBC tissues exhibit a significant increase of BOP1 expression, which promotes Wnt/β-catenin signaling [22]. Chemoresistant TNBC patients frequently have ST8SIA1 overexpression. Inhibition of ST8SIA1 improved chemotherapeutic effectiveness in TNBC cells and suppressed the FAK/Akt/mTOR and Wnt/β-catenin signaling pathways [23]. Also, In MDAMB231 cells, by activating the AF-6/ERK signaling pathway and upregulating CSC characteristics, CLDN6 promotes chemoresistance to ADM [24]. Twist upregulation activating the self-renewing factor Bmi1 and the epithelial-mesenchymal transition allows USP2 to maintain the CSC population, which is spared in chemotherapy and causes tumor recurrence and progression [25]. In TNBC, CHK1 is preferred to homologous recombination (HR) in avoiding replication stress. Errors in double-strand break repair can be compensated for by the ATR-CHK1 signaling cascade [26]. Rac1 activates aldolase A and ERK
signaling, causing the up-regulates glycolysis and especially the non-oxidative pentose phosphate pathway (PPP), leading to the increase of nucleotides metabolism which enables breast cancer cells to avoid chemotherapeutic-induced DNA damage [27]. By targeting the FN1 receptor, ITGA5, overexpression of miR-326 reduced FN1-driven chemoresistance [28]. By lowering p53-promoted miR-205-5p expression, TNFAIP8 was critical in the DDP tolerance formation of TNBC cells [29]. The chemoresistance and stemness of TNBC cells involve the expression of FAM83A, which can be regulated by miR-613 [30]. In addition, there is a novel oncogenic role of THEMIS2, and the underlying mechanism is that it suppresses PTP1B's association with MET, resulting in its activation [31]. By rewiring MAPK feedback and cross-talk, RASAL2 acts as a TNBC chemoresistance mediator and confers high collateral sensitivity to MEK1/2 and EGFR inhibitor combinations [32]. Notch1 promotes EMT and chemoresistance, together with invasion and proliferation of TNBC cells, by directly activating the MCAM promoter [33].

However, there are very few studies that systematically list a large number of chemoresistant markers in TNBC rather than focusing on one or few genes. This paper employed single-cell RNA-seq to find all chemoresistant markers in the TNBC samples.

3. Methods

This study employed Dataset GSE118389 and paper [7], 2018 from Simona Cristea from Dana-Farber Cancer Institute was published on Aug 11, 2018. It includes single-cell RNA sequencing of 1,534 cells in fresh tumors from six TNBC patients.

3.1. Data

According to the original study, the sample included tumors from six women who had invasive ductal carcinomas that were primary, nonmetastatic, and triple-negative before receiving any local or systemic treatment. The tumors had various degrees of immune and stromal infiltration as well as a dense mass of invasive ductal carcinoma cells, which are all histological features of TNBC. Out of the six tumors with sufficient tissue for analysis, two (tumors 84 and 126) have local axillary lymph node involvement.

The Dana Farber/Harvard Cancer Center Institutional Review Board approved the collection of fresh tumors from TNBC specimens at Massachusetts General Hospital (93-085).

The sample includes 868 epithelial cells, 94 stroma cells, 64 macrophages, 53 T cells, 19 B cells, 14 endothelial cells, 19 undecided cells, and 58 unknown cells. Expression markers used to categorize cell types was from different references and consists of 49 expression markers specific to four cell types. Then, t-SNE-based clustering on a projection of the cells into a lower dimensional space in Monocle was applied to assign cell types

Monocle was used to cluster the 868 epithelial cells, which employed a density-based approach to choose the number of clusters automatically. cluster 1 (22 cells); cluster 2 (398 cells); cluster 3 (231 cells); cluster 4 (170 cells); cluster 5 (47 cells) have been identified as epithelial clusters.

Sequencing of single-cell RNASeq (scRNA-seq) samples was performed. Using the reference genome version GRCh38, RSEM with default parameters were used to quantify FASTQ files to transcript per million (TPM) expression values.

To identify low-quality cells, the following metrics were used: Library size, Number of expressed genes, as well as Total amount of mRNA. There are 1, 326 cells that pass all three thresholds, with 5 of them being pooled samples.

To have a higher probability of keeping genes related to inter-patient heterogeneity, the genes identified as unexpressed were required to be unexpressed in each patient rather than across all patients. After filtering, 13, 280 genes remain from the original 21, 785. Then, the scRNA-seq data was normalized.

254
3.2. Data processing and subsetting
This study employed R package Seurat, which is designed for single-cell RNA-seq data quality control, analysis, and exploration, for data process and analysis. From single-cell transcriptome measurements, users will be able to merge several single-cell data types and identify and assess heterogeneity sources.

First, QC metrics were used to select cells for analysis. Cells that have unique feature counts over 8000 or less than 500 were filtered. After filtering, there are 1098 cells left. Second, the data was normalized with log normalization with a scale factor of 10000. Third, PCA and UMAP were run on the remaining cells. For PCA, 30 principal components were used to compute and store. For UMAP, the dimensional reduction used was PCA, and the dimensions to use as input feature was 1:20. 16 clusters were identified by UMAP. Fourth, the epithelial subpopulation was subsetted using markers provided by the Supplementary Methods from [7] with the threshold of average expression ≥ 2 in at least 2 markers or percent expressed > 50% in at least one of "EPCAM", "KRT8", "KRT18", "KRT19", and 695 epithelial cells were isolated [7]. In order to lower the false positive rate, the threshold was tuned as stated. Then, a differential expression test was run to find differentially expressed genes between basal epithelial cells and luminal epithelial cells, the only 2 subpopulations that can be clearly identified in the epithelial subpopulation.

3.3. Differential gene expression analysis
After, cluster 2 was identified in the epithelial subpopulation using markers provided by the Supplementary Methods from [7] with the threshold of average expression ≥ 1 in at least 3 markers from the top 10 differentially expressed genes characterizing cluster 2 [7]. The threshold was tuned as stated to lower the false positive rate. 261 cluster 2 cells are isolated. Then, a differential expression test was run to find differentially expressed genes between cluster 2 and other clusters in the epithelial subpopulation (cluster 1,3,4 and 5).

Next, the basal epithelial subpopulation, containing 51 cells was isolated and PCA and UMAP were run. But cluster 2 was not able to be identified using markers provided by the Supplementary Methods from [7].

Then, the luminal subpopulation, containing 101 cells, was isolated and PCA and UMAP were run. After, cluster 2 was identified using markers provided by the Supplementary Methods from [7], and a differential expression test was run to find differentially expressed genes between cluster 2 and other clusters.

Statistics were done using R 4.0.5 (R Core Team, 2021) and the Seurat (v4.0.0; Satija, 2021) packages.

4. Results
This study used the epithelial subpopulation from GSE118389, a dataset from single-cell RNA sequencing of cells of TNBC tumors [7]. The epithelial subpopulation contains 695 cells, of which 51 are basal epithelial cells and 101 are luminal epithelial cells.

As discovered in [7], there is one cluster (cluster 2) in epithelial cells that has expressions of various signatures of chemoresistance. Therefore, the epithelial subpopulation was isolated and interrogated to identify genetic markers for chemoresistance in TNBC.

4.1. Differentially expressed markers in various subpopulations
First, cluster 2 was identified using markers provided in [7] and was compared to the other four clusters. There are 17 differentially expressed genes in cluster 2 with a p-value less than 10^-40, namely EEF2, RPL3, CSRP1, EEF1A1, MYADM, RPL10, TUBA1A, AHNAK, RPLP0, SNURF, RPS3A, SRSF5, ITM2B, SNRPN, CCND1, CIRBP, CSNK1A1 (Fig. 1), which are all expressed more strongly in cluster 1,3,4 and 5. Also, there are 1117 differentially expressed genes in cluster 2 with a p-value less than 10^-10, and 1987 differentially expressed genes in cluster 2 with a p-value less than 10^-5.
(a) Dot plot of top 5 of 17 differentially expressed genes between cluster 2 and other clusters with a p-value less than $10^{-40}$ in the epithelial subpopulation.

(b) Dot plot of 6–10 of 17 differentially expressed genes between cluster 2 and other clusters with a p-value less than $10^{-40}$ in the epithelial subpopulation.

(c) Dot plot of 11–17 of 17 differentially expressed genes between cluster 2 and other clusters with a p-value less than $10^{-40}$ in the epithelial subpopulation.

**Figure 1.** Differentially expressed genes between cluster 2 and other clusters in epithelial subpopulation.

Next, the epithelial subpopulation was separated into basal epithelial subpopulation and luminal epithelial subpopulation to compare the difference in expression between cell types as they are the only 2 subpopulations that can be clearly identified in the epithelial subpopulation. There are 11 differentially
expressed genes with a p-value less than $10^{-20}$, namely TAGLN, MYLK, MATN2, ACTA2, KRT14, MYL9, ANXA1, C2orf40, PTN, PDGFA, KRT17 (Fig. 2), which are all expressed more strongly in the basal epithelial subpopulation, 101 differentially expressed genes with a p-value less than $10^{-10}$, and 301 differentially expressed genes with a p-value less than $10^{-5}$.

Figure 2. Differentially expressed genes between basal and luminal epithelial subpopulations.

Then, cluster 2 was separated from other clusters in the luminal epithelial subpopulation. There are 17 differentially expressed genes in cluster 2 with a p-value less than $10^{-5}$, namely CCDC74B, DHRS2, THRSP, PTHLH, TACSTD2, KLK11, RAB31, SERPINA1, CLIC6, GGH, RPL39, IFITM1, WFDC2, DDX5, RPLP1, CRIP1, NPY1R (Fig. 3). PTHLH, TACSTD2, WFDC2 and DDX5 are more strongly expressed in cluster 1,3,4 and 5, whereas all other 13 differentially expressed genes are more strongly expressed in cluster 2.
(a) Dot plot of top 5 of 17 differentially expressed genes between cluster 2 and other clusters with a p-value less than $10^{-5}$ in the luminal epithelial subpopulation.

(b) Dot plot of 6~10 of 17 differentially expressed genes between cluster 2 and other clusters with a p-value less than $10^{-5}$ in the luminal epithelial subpopulation.

(c) Dot plot of 11~17 of 17 differentially expressed genes between cluster 2 and other clusters with a p-value less than $10^{-5}$ in the luminal epithelial subpopulation.

**Figure 3.** Differentially expressed genes between cluster 2 and other clusters in luminal epithelial subpopulation.
However, cluster 2 cannot be isolated in the basal epithelial subpopulation. This suggests that the basal epithelial subpopulation behaves very differently from the luminal epithelial subpopulation.

Last, the differentially expressed genes in cluster 2 in the epithelial subpopulation and those in the luminal epithelial subpopulation were compared. There are 7 common differentially expressed genes with a p-value less than $10^{-5}$: CCDC74B, THRSP, PTHLH, TACSTD2, RAB31, SERPINA1, DDX5, RPLP1. There is some overlap between differentially expressed genes in the epithelial subpopulation and the luminal epithelial subpopulation.

5. Discussion
An analysis of epithelial subpopulations of TNBC samples from dataset GSE118389 revealed 1987 differentially expressed genetic markers between chemoresistant clusters and non-chemoresistant clusters, as well as 17 differentially expressed genes between luminal epithelial subpopulations of TNBC samples. There is a noticeable amount of overlap between these markers. The following is a brief discussion on a few significantly differentially expressed markers.

Through activating and deactivating, eEF2 mediates the translocation phase of protein translation, which is controlled by eEF2 kinase phosphorylation. Since protein translation requires a large portion of cellular energy, it appears that pre regulatory and regulatory activities in eukaryotic translation are controlled by various signaling pathways. In higher eukaryotes, eEF2 may be one of the crucial elements regulating protein synthesis when there is a lack of energy or nutrients [34]. Additionally, as eEF2K and autophagy are crucial for maintaining aggressive tumor behavior and chemoresistance in resistant TNBC, eEF2K silencing is a possible cutting-edge approach for TNBC treatment [35].

The essential cell cycle, apoptosis, and DNA repair regulator rpL3 regulates the expression of p21 in cells in response to chemotherapy. Also, rpL3 is capable of controlling DNA repair without being dependent on the cell's p21 state. Additionally, rpL3 silencing eliminates the cytotoxic effects of 5-FU and L-OH, demonstrating that the absence of rpL3 causes chemotherapeutic drugs to be ineffective [36].

Members of the cysteine-rich protein (CRP) family, which mediates protein-protein interactions and is crucial for cell differentiation, cytoskeletal remodeling, and transcriptional regulation in vertebrates, are characterized by two LIM domains linked to short glycine-rich repeats [37]. In colorectal cancer, CSRP1 was thought to be a tumor suppressor gene [38]. Furthermore, CSRP1 may be inactivated as a result of aberrant methylation, and it could be a useful diagnostic marker for liver cancer [39]. However, suppression of CSRP1 expression by celecoxib may exhibit anti-gastric cancer effects [40].

The results were compared to Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing and its dataset SRP114962 [41]. This dataset is from samples of various cell types. However, none of the differentially expressed genes in cluster 2 in either epithelial subpopulation or luminal epithelial subpopulation is identical to those in [41]. This suggests samples from different cell types result in very different differentially expressed genes.

There are several limitations in this analysis: First, the sample size is small. Only 695 epithelial cells from 6 TNBC tumors were included in the analysis. A larger sample size may be able to generate more accurate results. In addition, since the thresholds for the presence of markers were raised for subsetting epithelial and cluster 2 cells to lower the false positive rate, some cells may have been subsetted into other subsets, which may cause inaccuracy.

6. Conclusion
After an analysis of dataset GSE118389, 1987 genetic markers that were differentially expressed between a chemoresistant cluster and non-chemoresistant clusters in epithelial subpopulations of the TNBC samples were found, and 17 differentially expressed genes between a chemoresistant cluster and non-chemoresistant clusters in luminal epithelial subpopulations of the TNBC samples were found. There is a considerable overlap between these markers.

References


