# ADC Target Biomarker Discovery in the Hormone-sensitive Cancer Using Gene Expression Data

## Introduction:

The use of high-throughput sequencing techniques has made the identification of potential biomarkers possible. However, the challenge now is to identify biomarkers that can lead to more targeted treatments, such as those that can be applied exclusively to cancer cells without harming the normal healthy cells, in order to avoid the common side effects often seen with cancer treatments. Antibody-Drug Conjugate (ADC) therapy has gained a lot of attention in recent years with the goal of delivering targeted drugs exclusively to cancerous cells and avoiding the side-effects of cancer treatments. This study looks that previously published datasets for hormone-sensitive cancer types, and tries to identify potential cell surface receptor biomarkers that can be used with ADC.

## Background:

There has been a lot of research revolving around discovering the potential biomarkers for use with ADC drug therapy, with the focus being on identifying protein surface-membrane receptors that normally (and ideally) have low expression in normal cells, but high expression in cancerous/tumor cells [1]. ADCs consist of a monoclonal anti-body (mAb) that targets a protein surface receptor, ideally only on the cancer cell, as well as connecting linker, which is connected to a cytotoxic agent designed to kill the cancerous cell once it binds to the receptor and is taken up by the cell [2].

There are currently very few biomarkers that are approved by FDA for use with ADC. A few of the reasons include a lack of tumor response or excessive normal tissue toxicity in clinical trials [3]. This suggests that techniques need to be developed that can effectively differentiate the membrane-receptors found only on cancer cells, and although there are several papers that investigate this, more insight is needed.

Hormone sensitive cancers include breast cancer, Ovarian cancer, Uterine cancer as well as prostate cancer. These cancers are typically driven by sex hormones, including endogenous and exogenous hormones, the dysfunction of which leads to cancerous cell proliferation [4]. Anbarasu and Anbarasu, 2023, reviewed cancer biomarkers linked to the sex hormone receptors as well the recent therapeutic advancements. The authors specifically look at the molecular pathways that are affected by these sex hormone receptors in the different cancer types, examining biomarkers linked to sex hormone receptors. The authors suggest that these biomarkers can potentially be used as diagnostic and prognostic tools, as well as therapeutic targets [5].

The paper by Emmenis et al, 2024, discusses the investigation to identify novel cell surface receptor biomarkers in a more aggressive form of prostate cancer, the neuroendocrine prostate

cancer (NEPC). The research identified CELSR3 as a novel biomarker discovered via differential expression analysis, in the absence of an antibody to experimentally validate its presence in the different types of PC tissue. According to the study, CELSR3 is expressed in NEPC cancer cells and not in normal cells, and its knockdown resulted in reduced NEPC tumor cell proliferation and migration in vitro. They were able to develop a bispecific mAb to simultaneously bind to the CELSR3 on the tumor cell and the CD3 receptor present on human T cells, which resulted in CELSR3+ specific cell lysis. The research also identified other potential biomarkers, such as CEACAM5 and HMMR as well as transcript isoforms, that may also be potential targets [6]. This can lead to a more targeted treatment of cancer cells, and can be replicated for different types of cancers to look for biomarkers that are exclusive to those tumor cells.

Another paper by Gamble, P., Jaroensri, R., Wang, H. *et al.* determining breast cancer biomarkers using deep learning. The data collected is used to create deep learning models, such as convolutional neural networks and other advanced ML techniques to automatically analyze histopathology images and predict biomarker status based on the morphological features.

An extension of these techniques is poddibly the use of spatial transcriptomics (ST) data to quantify the mRNA population of a specific biomarker in the spatial context of intact tissue. The paper by Berglund, E., Maaskola, J., Schultz, N. *et al.*, 2018., reveals the heterogeneity of cancer subtypes within the prostate cancer tissue. Thus, integrating ADC analysis along with other techniques seems to be an effective way to understand the complexity of the Tumor Micro-Environment (TME), and use that to more precisely target the tumor receptor proteins that are found using that analysis.

Objectives:

The main objective of the study is to develop a pipeline to identify differentially expressed genes as potential biomarker candidates with low expression in normal cells and high expression in cancerous cells selected from mRNA expression data of various hormone-sensitive cancer types, and use that to determine potential candidates for ADC treatment.

- Select expression data derived from microarray or RNA-seq libraries that are available online, preferably from public databases such as GEO. I decided to look at the datasets from the study by Ahmadi, S., Sukprasert, P., Vegesna, R. *et al, 2022,* which uses gene expression data from different datasets to identify genes encoding cell surface receptors for use with various combinational therapies. I hope to identify specific receptor genes which may be targeted by ADC drugs. I will also be looking at multiple datasets from the different types of hormone-sensitive cancers, including GEO2R - GEO - NCBI (nih.gov),
- Identify biomarkers suitable for treatment with ADC using the datasets selected via differential expression analysis as well as machine learning techniques to develop a predictive model based on the gene expression signatures identified.

- Validate the identified biomarkers in pre-clinical models and patient cohorts via existing gene expression independent studies to endure their generalizability, as well as through survival plot validation
- Conduct functional enrichment analysis using tools such as Gene Set Enrichment Analysis (GSEA) to understand the biological pathways represented by the identified biomarkers, providing insight to the current mechanisms of the biomarkers and how they can potentially help with cancer treatment.

Methodology:

- Data collection: Acquire gene expression data from previously done studies on RNA-seq analysis on tumor samples. This was done using TCGA Firehose Legacy RNA-Seq count data. The raw and preprocessed data was downloaded and then analyzed using DESEQ2 to get the differentially upregulated genes. The second dataset used was from GEO with the ID: GSE158085. The list of all surface markers was derived from the paper by Emmenis et al, via their code repository on GitHub where the complete list of surface protein markers was provided. The single-cell analysis was done using the data provided on the Single-cell portal via The Broad Institute website. The mRNA expression comparison for the selected differentially upregulated genes between ER+ and ER- cells was done on the METABRIC dataset on cBioPortal.
- Data preprocessing: The gene expression data was normalized across the different samples, missing data were handled, outliers identified and managed, as well as correction for batch effects were done if needed.
- Differential Gene Expression Analysis: Use computational approaches such as differential expression analysis between normal and cancer cells, via R packages such as DESeq2, to identify potential biomarkers for ADC, as well as incorporate single-cell analysis techniques to characterize the distribution of the biomarker-expressing cell populations.
- Validation: Candidate biomarkers were validated via multiple datasets on patient cohorts from existing clinical outcomes data.
- Functional Enrichment Analysis: The biomarkers identified via differential expression as well as scRNA analysis were subjected to enrichment analysis via GSEA to help interpret the biological significance of the biomarkers.

Expected Outcome:

The primary expected outcome of this research is the identification of tumor cell surface receptors using the RNA-seq and differential expression analysis as potential biomarkers to use with ADC drug therapy for hormone-sensitive cancers like breast and prostate cancer. Insight into the organization of the single-cell populations in the TME and their distributions of the potential biomarkers could allow for a better understanding of the mechanisms that govern the TME. By using that information and identifying potential differentially expressed biomarkers, I hope to develop a means to identify ADC biomarkers.

This study can help enhance personalized medicine in several ways. Firstly, identifying potential ADC surface-protein biomarkers on cancer cells can allow for targeted treatment, which mitigates the side-effects associated with more generalized chemotherapy approaches. Secondly, creating classifier models can help determine if the identified biomarkers are indeed useful for ADC targeting.

## Results and Discussion:

Based on the differential analysis using DESEQ2 from TCGA and GEO GSE158085 datasets, I was able to identify 1699 genes that were upregulated in TCGA and 1859 genes that were upregulated in the GEO dataset, with 301 overlaps, as can be seen in the Venn diagram (Fig 1a). I then did an overlap of these genes with the combined surface protein list provided in the paper by Emmenis et al, Cancer Res Common, 2023. This provided 219 genes from the TCGA list of upregulated genes and 215 genes from the GEO list of upregulated genes. Overlapping these two lists further provided 43 genes (Fig 1b) that were found to be upregulated in both the datasets as well as were present in the surface protein markers list.

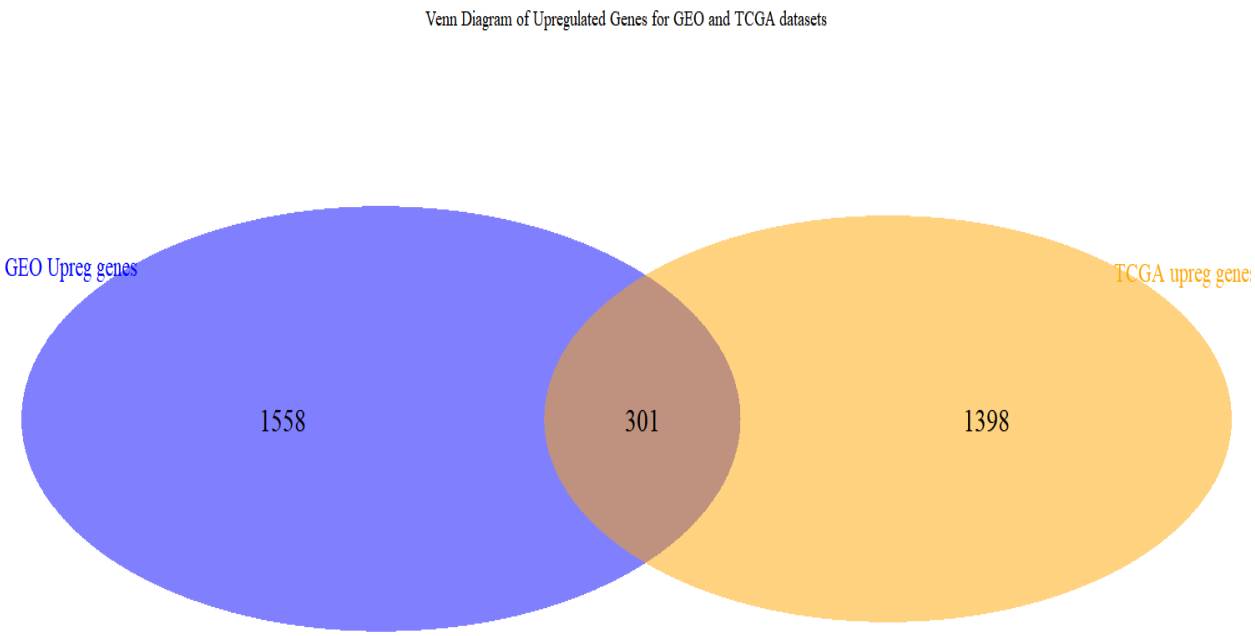


*Figure 1a*

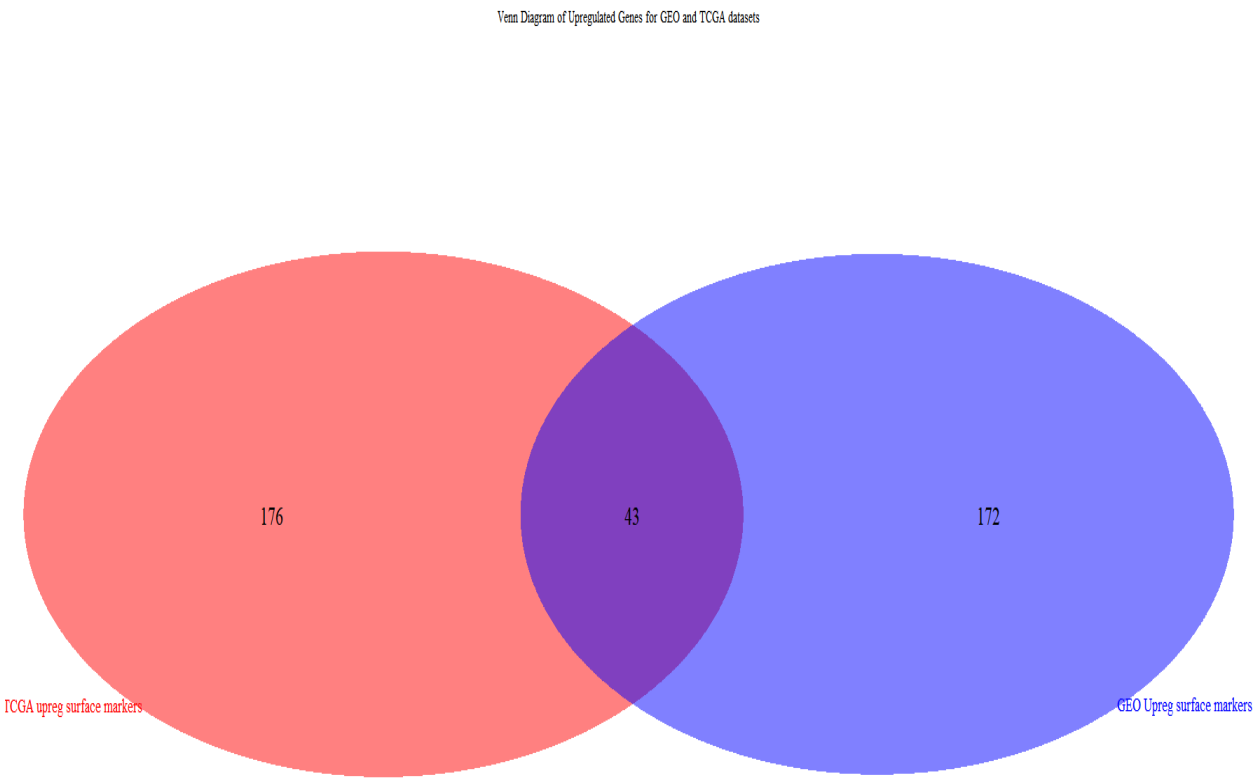


*Figure 1b*

An enrichment analysis of these two datasets via GSEA also provided the list of pathways that were significantly upregulated and downregulated in these two datasets. A heatmap of the NES values was created to compare the significant pathways in these two datasets, with several overlaps, as can be seen in Fig 2.

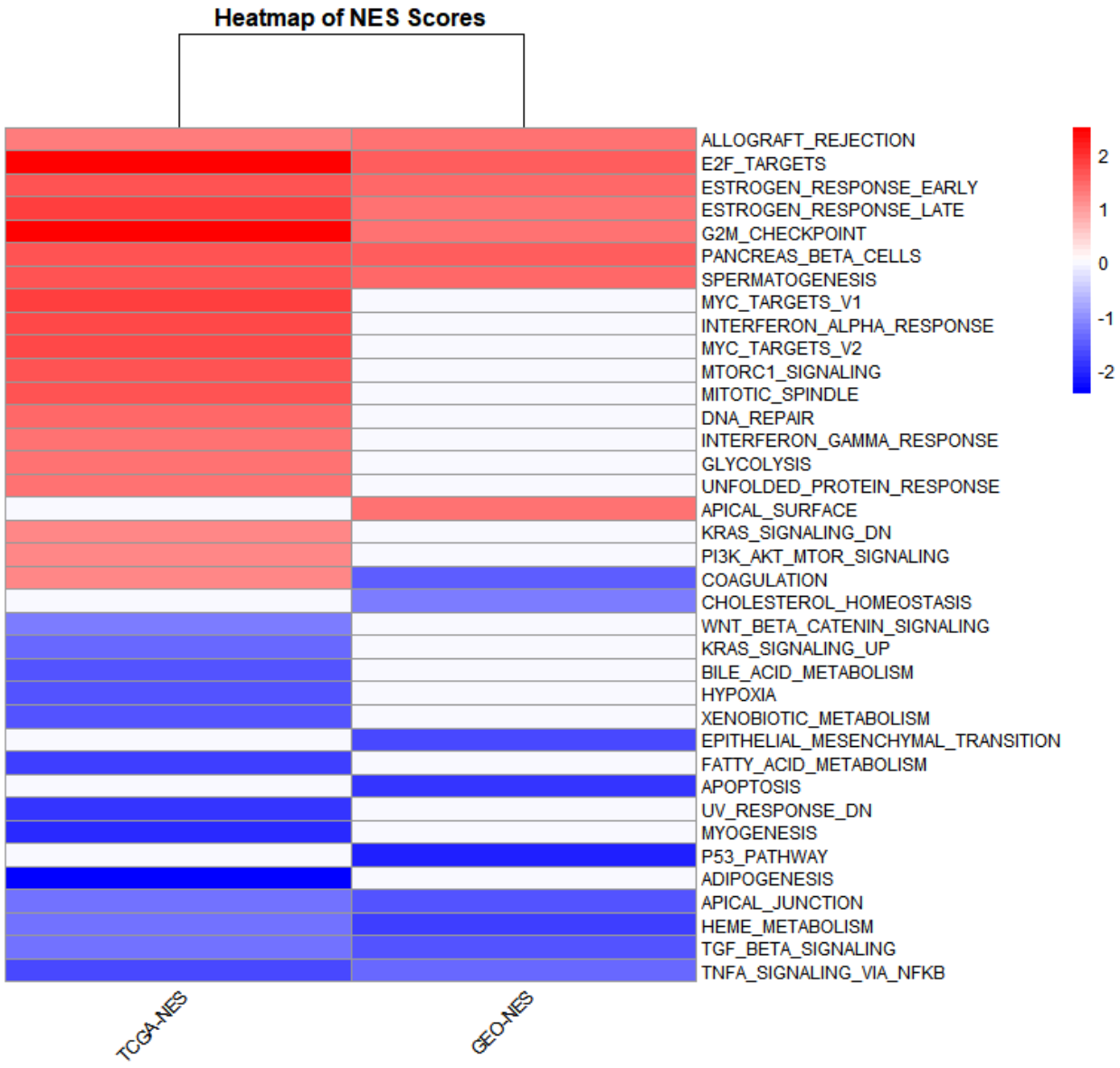


*Figure 2*

I was also curious to see if the genes identified were more highly expressed in ER+ cancer types or in the triple negative cancers, or both. For this I used the METABRIC dataset on breast cancer that is available publicly online on cBioPortal. Separating the dataset based on their 3-Gene classifier subtype, and using the two ER+/HER- low proliferation and ER-/HER- subtypes allowed me to create gene expression boxplots for each of the genes, showing whether the difference in expression between the two categories was significant or not, as well as which category was it upregulated for. This information is provided in Fig. 3, for certain genes.

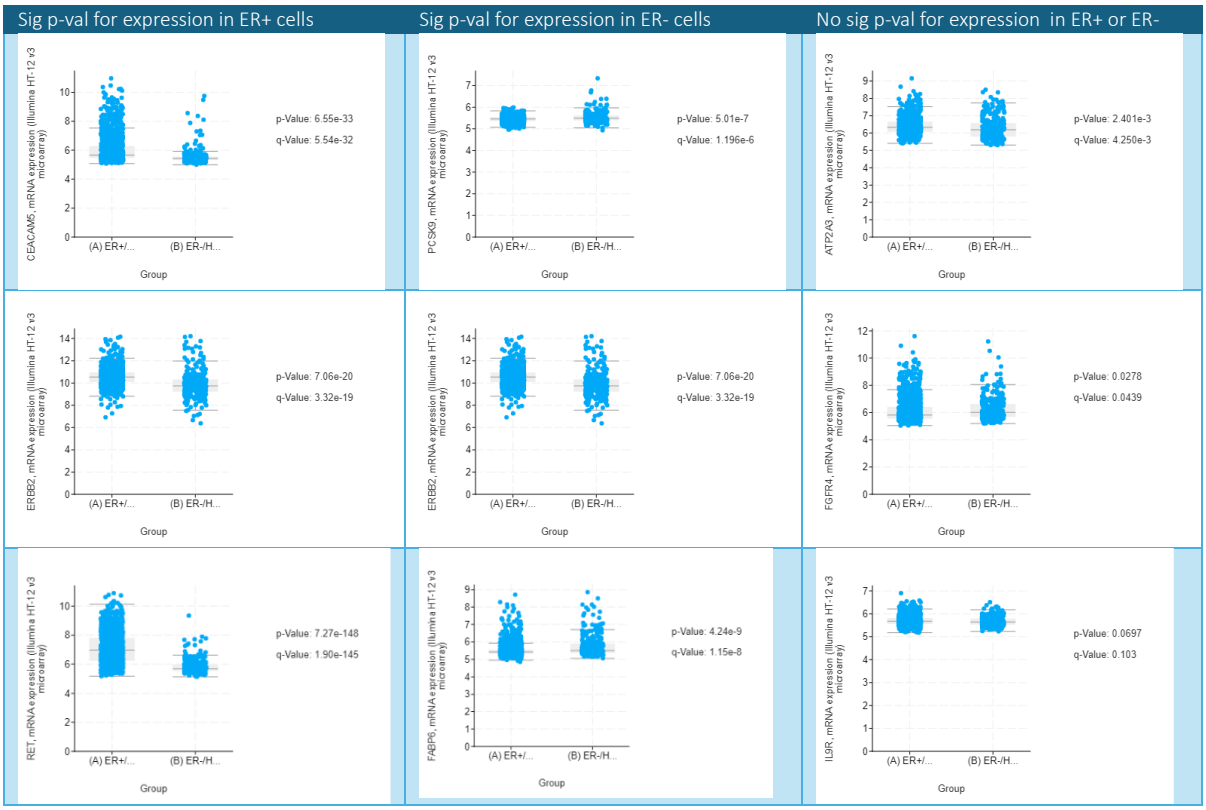


*Figure 3. Box plots for significant p-vals for ER+ (1st row), ER- (2nd row) genes and for no significant p-values between ER+ and ER- expression in patients (3rd row)*

I then conducted a single-cell RNA-Seq expression analysis for the 43 genes identified above using the whole BRCA miniatlas dataset provided by the single-cell portal of the Broad institute [10]. The results of the expression for these genes is individually shown in the violin plots, showing their expression in each of the major cell types included in the study. The violin plots were modified such that if the expression in any cell type category was lower than 10% of the entire datapoints for that gene, the data for that cell type was not included in the violin plot. Some examples of the violin plots can be seen in Fig 4. I also created similar violin plots for the House-keeping genes (HKG) in order to compare the expression of these genes with the expression of the upregulated surface protein genes identified, so that the genes with very low expression compared to the standard can be flagged.

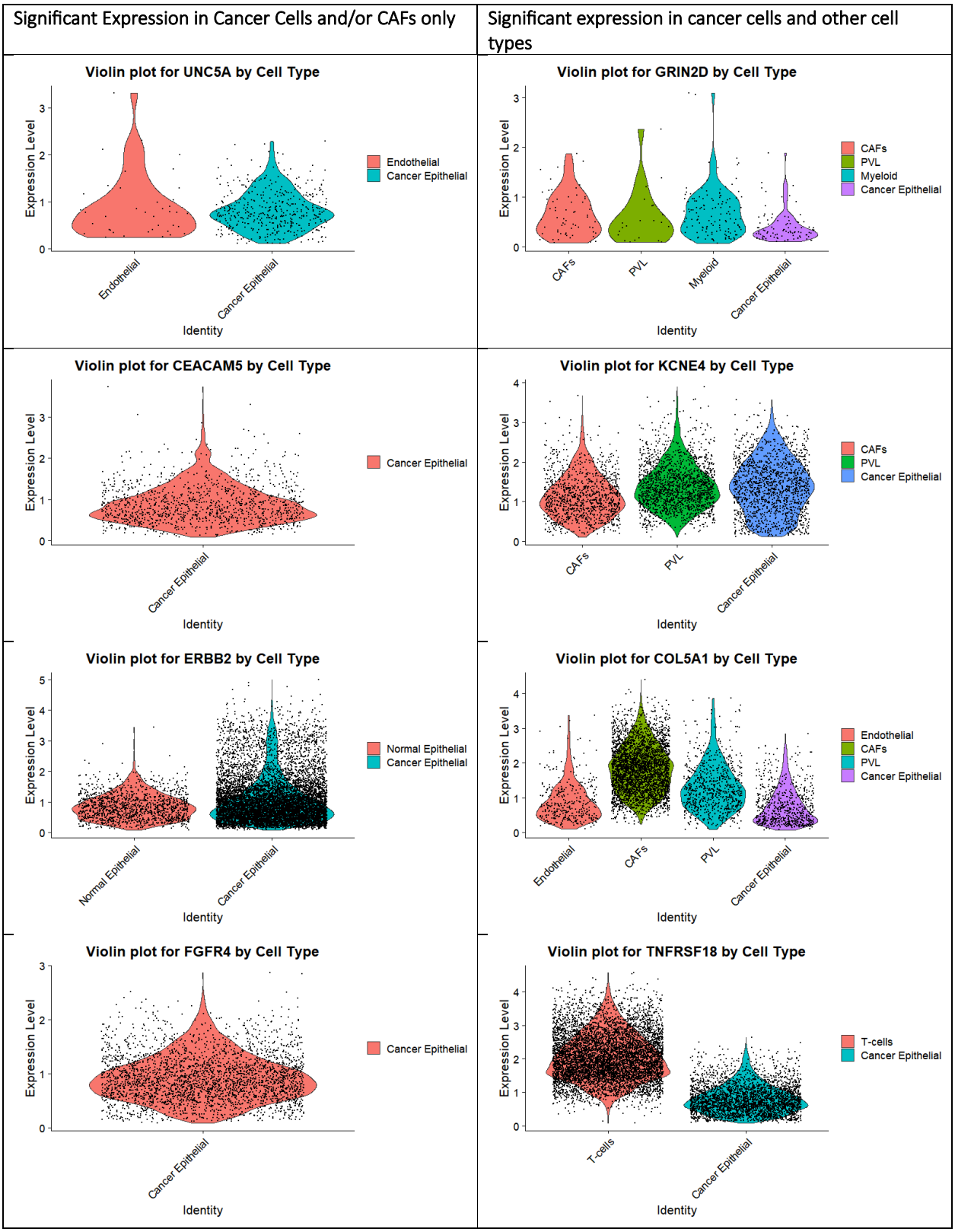


*Figure 4: Single-cell RNA-Seq analysis of some upregulated genes shown here*

Finally, I created a summary heatmap of all the findings, as shown in Fig 5, including information on which of the upregulated genes identified in both datasets were surface markers, ER+, were only expressed in cancer epithelium or CAFs, and whether there were approved drugs already available for them.

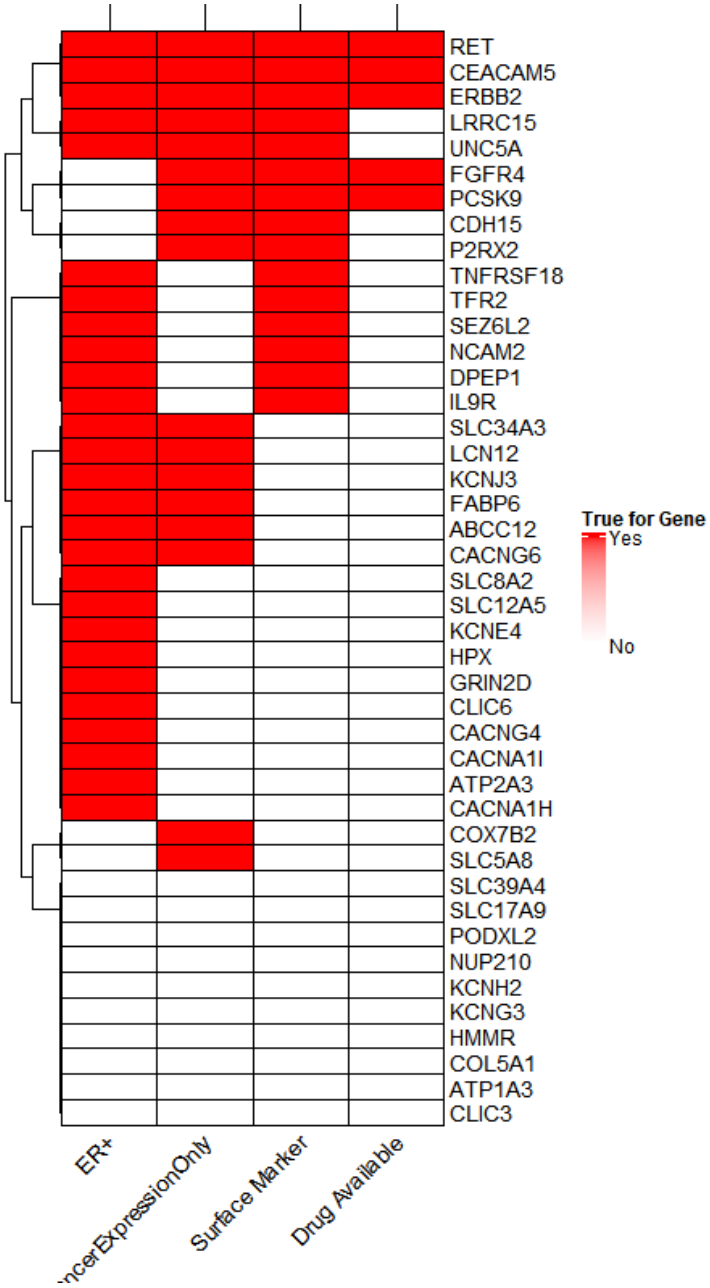


*Figure 5: Summary Heatmap of Findings*

Based on the results, out of the 1699 genes that were upregulated in TCGA and 1859 genes that were upregulated in the GEO dataset, there were only 301 overlaps of the upregulated genes. This was an interesting finding because both these datasets looked at breast cancer patients and compared their expression to 'normal' healthy individuals' expression.

Out of the 301 overlapping genes, only 43 were those of surface markers or related proteins as identified via the list provided by Emmenis et al. Out of these 43 proteins identified, only 17 had expression confined to cancer cells or CAFS, and only 15 were actual cell surface receptor that could potentially be used with ADCs. 27 of the genes identified were highly expressed for ER+. In terms of targeted drugs for these proteins, currently they are available for only 5 out of the 15 surface receptor genes identified. These 5 genes are RET, CEACAM5, ERBB2, FGFR4 and PCSK9, and as can be seen in figure 6, these have differential expression only in cancer cell based on the single cell analysis and are cell surface receptor, although FGFR4 and PCSK9 are expressed in ER-negative cancer types. The information on the genes which have drugs currently available for them is provided in table 1.

Table 1

| Surface Marker | Found in Hormone sensitive cancers | Drugs Available | Other applications |
|---|---|---|---|
| RET | Found in aa subset of ER+ breast cancers | RET inhibitors (e.g. selpercatinib, pralsetinib) | Multi-kinase inhibitors are effective in RET fusion positive cancers, such as NSCL and medullary thyroid cancer |
| CEACAM5 | More commonly found in prostate cancer | ADCs e.g. Labetuzumab-SN-38 | Also found and treated in colorectal cancer |
| ERBB2 | Well-established target in breast cancer | Monoclonal antibodies (e.g. trastuzumab and pertuzumab); ADCs (trastuzumab emtansine); and small-molecule inhibitors (lapatinib, neratinib) | |
| FGFR4 | Linked to poor outcomes in breast cancer | Futibatinib and other FGFR inhibitors being evaluated for clinical trials | |
| PCSK9 | Being explored for secondary tumor-supportive roles | Targeted currently by inhibitors like Alirocumab (Praluent) and Evolocumab (Repatha) for lowering LDL cholesterol levels | Lowering LDL cholesterol levels |

For example, CEACAM5 is primarily involved in cell adhesion, signaling and immune response modulation. As a member of the Ig superfamily, CEACAM5 is known to promote tumorigenesis by inhibiting anoikis [6]. In cancer cells it loses its apical polarity with increased expression and thus can be used as a good marker for cancer-related CEA expression. As such, CEACAM5 is found to be highly expressed on the cancer cell surfaces of some epithelial tumors, including colon, gastric, pancreatic, ovarian and lung cancers.

According to the paper by Zhang et al, J.  of international Medical Research, 2020, the potential association of CEACAM5 with the p58 kinase, thereby inhibiting p58 activity and leading to tumorigenesis, was examined. The researcher deduced that CEACAM5 induces NSCLC tumor growth via the p38/SMAD2/3 pathway by in vivo experiments that showed higher p38 activity when CEACAM5 expression was downregulated in tumor cells. Similarly, overexpression of CECAM5 inhibited p38 activity and promoted invasion and metastasis [13].

In another research paper by DeLucia et al, Clin Cancer Res, 2021, examining CEACAM5 levels in neuroendocrine prostate cancer (NEPC), which is a lethal form of PC, the researchers found enriched CEACAM5 expression in NEPC compared with other metastatic CRPC subtypes. The researchers found a correlation between the pioneer transcription factor achaete-scute homolog

1 (ASCL1) expression and CEACAM5 and determined that ASCL1 can drive neuroendocrine reprogramming of prostate cancer via epigenetic regulation which is associated with increased chromatin accessibility of the CEACAM5 core promotor and hence increased CECAM5 expression [14].

The findings from the analysis point to the complex nature of genetic research. Although there were over 1500 differentially upregulated genes in each of the datasets analyzed, only ~ 20% of them overlapped, regardless of the fact that the samples from both datasets were taken from breast cancer tissue and compared against normal tissue. There could be several factors accounting for the disparity, including the type of breast cancer tissue, the patient sub-populations included, as well as the variability in the assay results.

Overall, this analysis revealed the top overlapping genes that are highly expressed in hormone-sensitive cancers such as breast and prostate cancer from multiple datasets, as well as analyzes the expression of these genes in several cell types from a single cell dataset. I also looked at the expression of these genes in ER+ and Triple-negative breast cancer samples, as well as examined which gene proteins currently have drugs developed for them. Finally, a pathway enrichment analysis was also done using GSEA for the gene list, and important gene pathways were identified which are consistent with current similar analyses.

References:

A list of references will be included as used for the background and methodology for this research.

Current citations:

[1] Razzaghdoust A, Rahmatizadeh S, Mofid B, Muhammadnejad S, Parvin M, Torbati PM, Basiri A. Data-Driven Discovery of Molecular Targets for Antibody-Drug Conjugates in Cancer Treatment. Biomed Res Int. 2021 Jan 2;2021:2670573. doi: 10.1155/2021/2670573. PMID: 33490264; PMCID: PMC7801065.

[2] Mythili Shastry et al., Rise of Antibody-Drug Conjugates: The Present and Future. *Am Soc Clin Oncol Educ Book* **43**, e390094(2023), DOI:10.1200/EDBK_390094

[3] Abdollahpour-Alitappeh M, Lotfinia M, Gharibi T, Mardaneh J, Farhadihosseinabadi B, Larki P, Faghfourian B, Sepehr KS, Abbaszadeh-Goudarzi K, Abbaszadeh-Goudarzi G, Johari B, Zali MR, Bagheri N. Antibody-drug conjugates (ADCs) for cancer therapy: Strategies, challenges, and successes. J Cell Physiol. 2019 May;234(5):5628-5642. doi: 10.1002/jcp.27419. Epub 2018 Nov 27. PMID: 30478951.

[4] Shinjini Ganguly, Divya Naik, Andrew Muskara, Omar Y Mian, The Nexus of Endocrine Signaling and Cancer: How Steroid Hormones Influence Genomic Stability, *Endocrinology*, Volume 162, Issue 1, January 2021, bqaa177, https://doi.org/10.1210/endocr/bqaa177

[5] Anbarasu, S., Anbarasu, A. Cancer-biomarkers associated with sex hormone receptors and recent therapeutic advancements: a comprehensive review. *Med Oncol* **40**, 171 (2023). https://doi.org/10.1007/s12032-023-02044-3

[6] Van Emmenis L, Ku SY, Gayvert K, Branch JR, Brady NJ, Basu S, Russell M, Cyrta J, Vosoughi A, Sailer V, Alnajar H, Dardenne E, Koumis E, Puca L, Robinson BD, Feldkamp MD, Winkis A, Majewski N, Rupnow B, Gottardis MM, Elemento O, Rubin MA, Beltran H, Rickman DS. The Identification of CELSR3 and Other Potential Cell Surface Targets in Neuroendocrine Prostate Cancer. Cancer Res Commun. 2023 Aug 3;3(8):1447-1459. doi: 10.1158/2767-9764.CRC-22-0491. PMID: 37546702; PMCID: PMC10401480.

[7] Berglund, E., Maaskola, J., Schultz, N. *et al.* Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nat Commun* **9**, 2419 (2018). https://doi.org/10.1038/s41467-018-04724-5

[8] Ahmadi, S., Sukprasert, P., Vegesna, R. *et al.* The landscape of receptor-mediated precision cancer combination therapy via a single-cell perspective. *Nat Commun* **13**, 1613 (2022). https://doi.org/10.1038/s41467-022-29154-2

[9] Gamble, P., Jaroensri, R., Wang, H. *et al.* Determining breast cancer biomarker status and associated morphological features using deep learning. *Commun Med* **1**, 14 (2021).

[10] Wu, S.Z., Al-Eryani, G., Roden, D.L. *et al.* A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet* **53,** 1334–1347 (2021). https://doi.org/10.1038/s41588-021-00911-1

[11] Yu, X., Meng, F. *et al.* Integrating spatial and single-cell transcriptomics reveals tumor heterogeneity and intercellular networks in colorectal cancer. *Cell Death Dis* **15**, 326 (2024). https://doi.org/10.1038/s41419-024-06598-6

[12] Nigro et al, RET in breast cancer: pathogenic implications and mechanisms of drug resistance, Cancer Drug Resist, 2019, **10.20517/cdr.2019.66**

[13] Zhang X, Han X, Zuo P, Zhang X, Xu H. CEACAM5 stimulates the progression of non-small-cell lung cancer by promoting cell proliferation and migration. Journal of International Medical Research. 2020;48(9). doi:10.1177/0300060520959478

[14] DeLucia et al, Regulation of CEACAM5 and Therapeutic Efficacy of an Anti-CEACAM5–SN38 Antibody–drug Conjugate in Neuroendocrine Prostate Cancer, Clin Cancer Res (2021) 27 (3): 759–774.https://doi.org/10.1158/1078-0432.CCR-20-3396

[15] Alexey Rak, Anand Kumar, Francis Duffi et al. Structural insights into epitope-paratope interactions of monoclonal antibody targeting CEACAM5-expressing tumors, 11 August 2023, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-3235785/v1]