





Article

Single-Cell Transcriptional and Epigenetic Profiles of Male Breast Cancer Nominate Salient Cancer-Specific Enhancers

Hyunsoo Kim ^{1,†}, Kamila Wisniewska ^{1,†}, Matthew J. Regner ^{1,2}, Aatish Thennavan ^{1,3}, Philip M. Spanheimer ^{1,4} 
and Hector L. Franco ^{1,2,5,*} 

- ¹ Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
² Bioinformatics and Computational Biology Graduate Program, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
³ Oral and Craniofacial Biomedicine Program, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
⁴ Division of Surgical Oncology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
⁵ Department of Genetics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
* Correspondence: hfranco@med.unc.edu
† These authors contributed equally to this work.

Abstract: Male breast cancer represents about 1% of all breast cancer diagnoses and, although there are some similarities between male and female breast cancer, the paucity of data available on male breast cancer makes it difficult to establish targeted therapies. To date, most male breast cancers (MBCs) are treated according to protocols established for female breast cancer (FBC). Thus, defining the transcriptional and epigenetic landscape of MBC with improved resolution is critical for developing better avenues for therapeutic intervention. In this study, we present matched transcriptional (scRNA-seq) and epigenetic (scATAC-seq) profiles at single-cell resolution of two treatment naïve MBC tumors processed immediately after surgical resection. These data enable the detection of differentially expressed genes between male and female breast tumors across immune, stromal, and malignant cell types, to highlight several genes that may have therapeutic implications. Notably, *MYC* target genes and *mTORC1* signaling genes were significantly upregulated in the malignant cells of MBC compared to the female counterparts. To understand how the regulatory landscape of MBC gives rise to these male-specific gene expression patterns, we leveraged the scATAC-seq data to systematically link changes in chromatin accessibility to changes in gene expression within each cell type. We observed cancer-specific rewiring of several salient enhancers and posit that these enhancers have a higher regulatory load than lineage-specific enhancers. We highlight two examples of previously unannotated cancer-cell-specific enhancers of *ANXA2* and *PRDX4* gene expression and show evidence for super-enhancer regulation of *LAMB3* and *CD47* in male breast cancer cells. Overall, this dataset annotates clinically relevant regulatory networks in male breast tumors, providing a useful resource that expands our current understanding of the gene expression programs that underlie the biology of MBC.

Keywords: breast cancer; male breast cancer; single-cell genomics; scRNA-seq; scATAC-seq; intratumoral heterogeneity; gene regulation; chromatin accessibility; enhancer elements



Citation: Kim, H.; Wisniewska, K.; Regner, M.J.; Thennavan, A.; Spanheimer, P.M.; Franco, H.L. Single-Cell Transcriptional and Epigenetic Profiles of Male Breast Cancer Nominate Salient Cancer-Specific Enhancers. *Int. J. Mol. Sci.* **2023**, *24*, 13053. <https://doi.org/10.3390/ijms241713053>

Academic Editors: Nicoletta Cordani and Monica Iorfida

Received: 21 July 2023

Revised: 18 August 2023

Accepted: 19 August 2023

Published: 22 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Male breast cancer (MBC) is a rare type of cancer that occurs in the breast tissue of men. MBC accounts for only 1% of total breast cancer incidence [1,2]; however, men are likely to present with larger, higher-grade tumors and more lymph node involvement compared to females with breast cancer [3]. Moreover, the 5-year mortality of MBC is higher than that of female breast cancer (FBC) [4], which may be due to the older age

at the time of diagnoses, delays in diagnosis, the presence of comorbidities, or intrinsic biological differences [3–5]. Most men with breast cancer are diagnosed with invasive ductal carcinoma, and their tumors are estrogen-receptor (ER) positive, progesterone-receptor (PR) positive, and HER2 negative [5–8]. At the molecular level, most men will present with Luminal A-like or Luminal B-like tumors [6,9], with some studies suggesting that MBCs have unique subtypes, M1 and M2, that differ from the intrinsic subtypes of FBC [10,11]. As with FBC, MBC risk increases for men with familial history of *BRCA* mutations. However, men show higher risk for *BRCA2* mutations than *BRCA1* mutations compared to women [12,13]. Men are also at higher risk of developing breast cancer if they are African American [14] or have comorbidities such as Klinefelter’s syndrome, hormone imbalances, liver disease, and obesity, among others [5,15,16].

Although MBC presents with similarities to certain FBC, the paucity of data related to the treatment of males makes it difficult to find targeted therapies [17–19]. Instead, men with breast cancer are treated in accordance with treatment paradigms for women, even if the efficacy of these treatments is low [18,20]. Interestingly, studies comparing the genomic profiles of male and female breast cancers have found important differences that are potentially driving outcomes of their respective disease [10,11,21,22]. For example, several microarray-based studies have found that *NAT1* [10], *mTOR* [23], *EIF4E* [23], *THY1* [11], and *SPAG5* [11] are upregulated in MBC compared to FBC and may serve as prognostic biomarkers. These initial studies highlight the fundamental differences between MBC and FBC and provide the impetus for defining male-specific disease mechanisms that will lead to better treatment options.

Single-cell genomics have revolutionized our ability to investigate the cellular, transcriptional, and epigenetic heterogeneity of human tumors with improved resolution. Single-cell RNA-seq (scRNA-seq) [24–29] refines our ability to measure the transcriptional profiles of thousands of individual cells within a particular tumor specimen to make conclusions about the underlying cellular heterogeneity of tumors and pinpoint salient gene expression programs. These gene expression programs are controlled and sustained by regulatory elements (e.g., *cis*-acting enhancer elements) scattered throughout the genome that are often re-wired and repurposed by cancer cells to drive oncogenic transcriptional programs [30–33]. The chromatin accessibility landscape may now be robustly profiled thanks to recent improvements in single-cell sequencing assay for transposase-accessible chromatin (scATAC-seq), revealing several layers of gene regulation, including *cis*-regulatory elements [34,35]. Together with scRNA-seq, scATAC-seq offers unprecedented resolution to reveal novel gene regulatory mechanisms in MBC. Although there are some notable cancer datasets with matched scRNA-seq and scATAC-seq [36,37], none have been reported for human MBC. Herein, we present matched transcriptional (scRNA-seq) and epigenetic (scATAC-seq) profiles of two treatment naïve MBC tumors processed immediately after surgical resection. First, we defined the differentially expressed genes between male and female breast tumors, specifically within the malignant cell types, to highlight several genes that may have therapeutic implications. Then, we systematically linked changes in chromatin accessibility to changes in gene expression to annotate the enhancer landscape within each cell type. Finally, we highlight the cancer-specific rewiring of several salient enhancers that drive abnormally high levels of male-specific gene expression and investigate the transcription factor occupancy at enhancers. Together, these data enable the annotation of the cellular composition, transcriptional, and epigenetic landscape of male breast tumors to help pinpoint drivers of this rare disease.

2. Results

2.1. Matched scRNA-seq and scATAC-seq of Male Breast Cancer

Two treatment-naïve MBC patients underwent mastectomy with curative intent (Figure 1). Immediately following surgical resection, each tumor was gently dissociated into a suspension of live cells using a gentle collagenase and hyaluronidase digestion and prepped for lipid droplet-based scRNA-seq and scATAC-seq via the 10× Genomics

Chromium system (Figure 1A). Each tumor specimen was divided into two pools to generate independent scRNA-seq and scATAC-seq libraries. Since these tumor specimens were never frozen or fixed in any way, a high level of cell viability during the dissociation process was maintained for robust sequencing coverage in single cells.

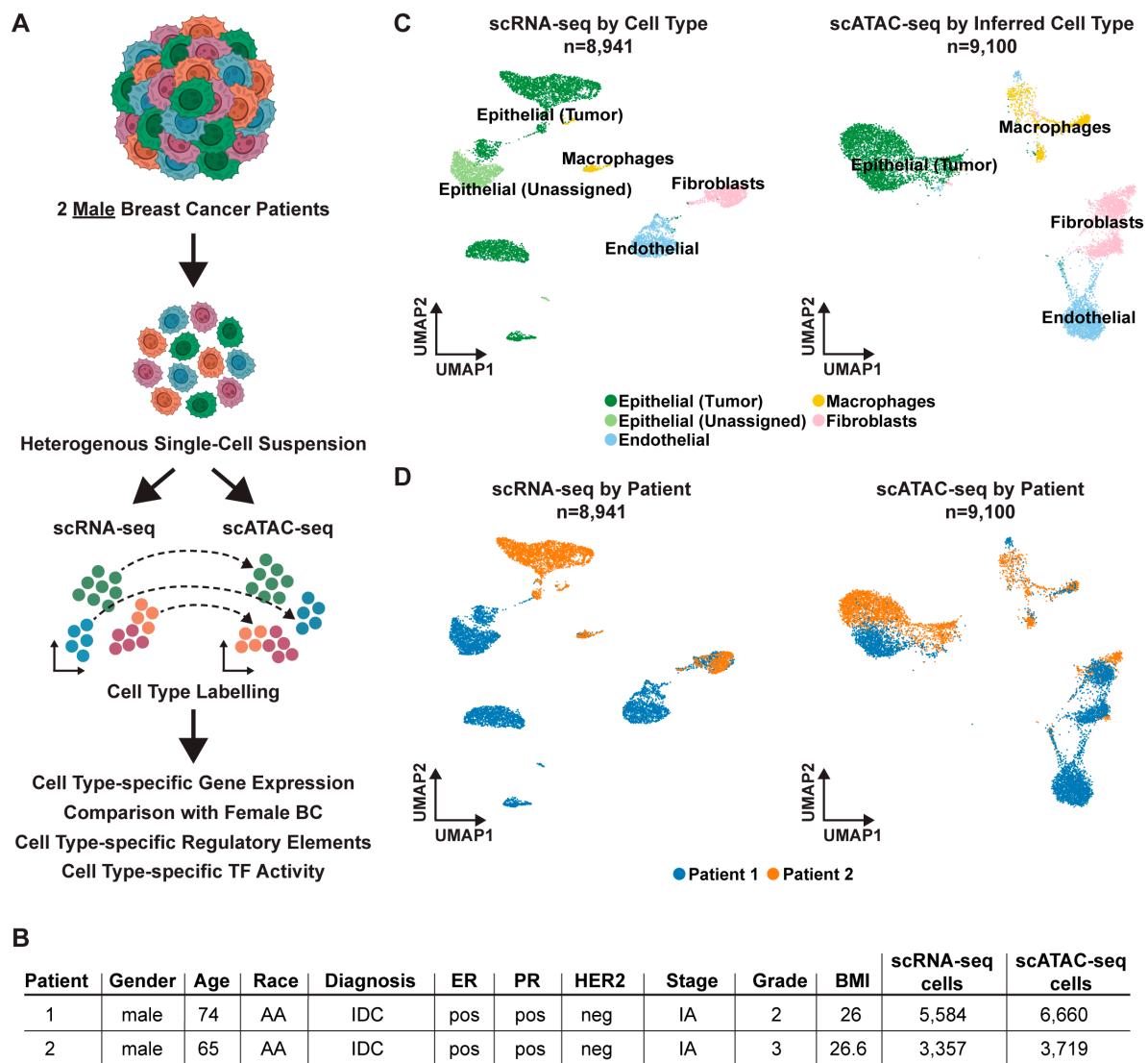


Figure 1. Overview of scRNA-seq and scATAC-seq of Male Breast Cancer Tumors. (A) Cartoon of scRNA-seq and scATAC-seq workflow. (B) Clinical features of the male breast tumors and total cells captured for both scRNA-seq and scATAC-seq. Abbreviations: AA—African American, IDC—Infiltrating Ductal Carcinoma, pos—positive immunohistochemical staining, neg—negative for immunohistochemical staining. (C) UMAP plot of scRNA-seq colored by cell type (left) and scATAC-seq colored by inferred cell type (right) across both male breast cancer patients capturing total cells after QC (see methods). (D) UMAP plot of scRNA-seq (left) and scATAC-seq (right) colored by patient of origin.

After quality control and doublet removal for each patient dataset, we obtained 8941 total cells profiled by scRNA-seq (Supplementary Figure S1A) and 10,379 total cells profiled by scATAC-seq (Supplementary Figure S1E). To ensure the robustness of our downstream analysis, we removed any cell type clusters that had less than 5000 RNA or ATAC fragment counts on average, or that could not confidently be assigned a cell type label (see Methods; Supplementary Figure S1C,G). After applying these filters, we

proceeded with a dataset composed of 8941 profiled by scRNA-seq cells and 9100 total cells profiled by scATAC-seq (Figure 1B).

To annotate the transcriptional profiles of the various cell types within these MBC tumor specimens, we performed principal component analysis (PCA) using the top 2000 most variable genes across all 8941 scRNA-seq cells. To correct for any technical variation, Harmony batch correction [38] was applied to the dataset and the cells were then classified into transcriptionally distinct clusters with graph-based clustering using the top 30 PCs and visualized using a Uniform Manifold Approximation and Projection (UMAP) plot. We identified four major cell types (epithelial cells, macrophages, fibroblasts, and endothelial cells) (Figure 1B, left) across nine clusters (Supplementary Figure S1A). Notably, the non-malignant cell types, such as the macrophages and fibroblasts, showed intermixing of cells from both patients, suggesting that these cell types had similar transcriptional profiles across patients. Conversely, the epithelial cell types (which were later confirmed to be malignant via inferCNV) showed patient-specific transcriptional profiles (Figure 1C, left). Because these data were batch corrected prior to clustering, we believe that the patient-specific transcriptional profiles of the epithelial cell types are biologically meaningful.

To analyze the chromatin accessibility landscape (vis-a-vis the epigenetic landscape) of these tumors, the scATAC-seq data was processed by creating a matrix of contiguous genomic tiles across the genome, in which we quantified Tn5 insertion counts across every cell. Then, we performed iterative latent semantic indexing (LSI) on the top 25,000 most variable genomic tiles [37,39]. We used Seurat v4.0.5 cross-modality integration [40] with the top 30 LSI dimensions (constrained to cells of the same patient tumor) to assign cell type labels from the matching scRNA-seq data to the scATAC-seq cells, and visualized the cells in the UMAP plot (Figure 1B, right) [41]. The scATAC-seq cells were clustered mostly by cell type and not by patient, which showed the quality of the dataset and data processing pipeline (Figure 1C, right). All four cell types (epithelial cells, macrophages, fibroblasts, and endothelial cells) were also observed in scATAC-seq (Figure 1B, right). Similar to the scRNA-seq, we found that the cell type clusters of macrophages and fibroblasts contained cells from both patients, while the cell type clusters of epithelial cells were patient-specific in scATAC-seq (Figure 1C, right). These observations likely reflect the biological overlap of the non-malignant cells across all patients and highlights the unique, and possibly tractable, biological features of the malignant cells within each patient's tumor.

2.2. Transcriptional Profiles of Male Versus Female ER+ Malignant Epithelial Cells

To determine the transcriptional differences between male and female cancer cells, we first identified the malignant epithelial cell types within each patient by inferring putative copy number events for each cell cluster using the inferCNV [42] approach (Supplementary Figure S2A–C). Then, we compared the cancer-epithelial cell profiles of our MBC tumors to publicly available scRNA-seq profiles of FBC epithelial cells [43] of the same molecular subtype. The male epithelial cells from both patients grouped together in a UMAP that contained both male and female cancer-epithelial cells, suggesting they have distinct transcriptional profiles compared to the female epithelial cells (Figure 2A). Differential gene expression analysis between MBC and FBC, at a single-cell resolution, identified 1004 upregulated genes and 14 downregulated genes in MBC with $\log_2FC > 0.25$ and adjusted p -value < 0.01 (Supplementary Table S1). To simplify the heatmap visualization, we filtered the differentially expressed genes with an adjusted p -value $< 1 \times 10^{-12}$ to arrive at the top 25 upregulated genes and top 5 downregulated genes (Figure 2B). To validate the soundness of the differential gene expression analysis, we visualized the expression of the top two upregulated genes (*RPS4Y1* and *TNFRSF12A*) and top two downregulated genes (*FOS* and *XIST*) in the UMAP plot (Figure 2C). Sex-specific genes such as *RPS4Y1* (located on the Y-chromosome) and *XIST* (needed for X-chromosome inactivation [44]) were exclusively expressed in the corresponding male and female epithelial cells, giving confidence to our analysis (Figure 2C). Gene ontology analysis [45] of the differentially expressed genes showed that the upregulated genes in MBC are enriched in 'MYC targets' and 'mTORC1

signaling', which suggests the possibility of using BET bromodomain inhibitors [46] and/or mTOR inhibitors [47,48] for MBC treatment (Figure 2D, Supplementary Table S1).

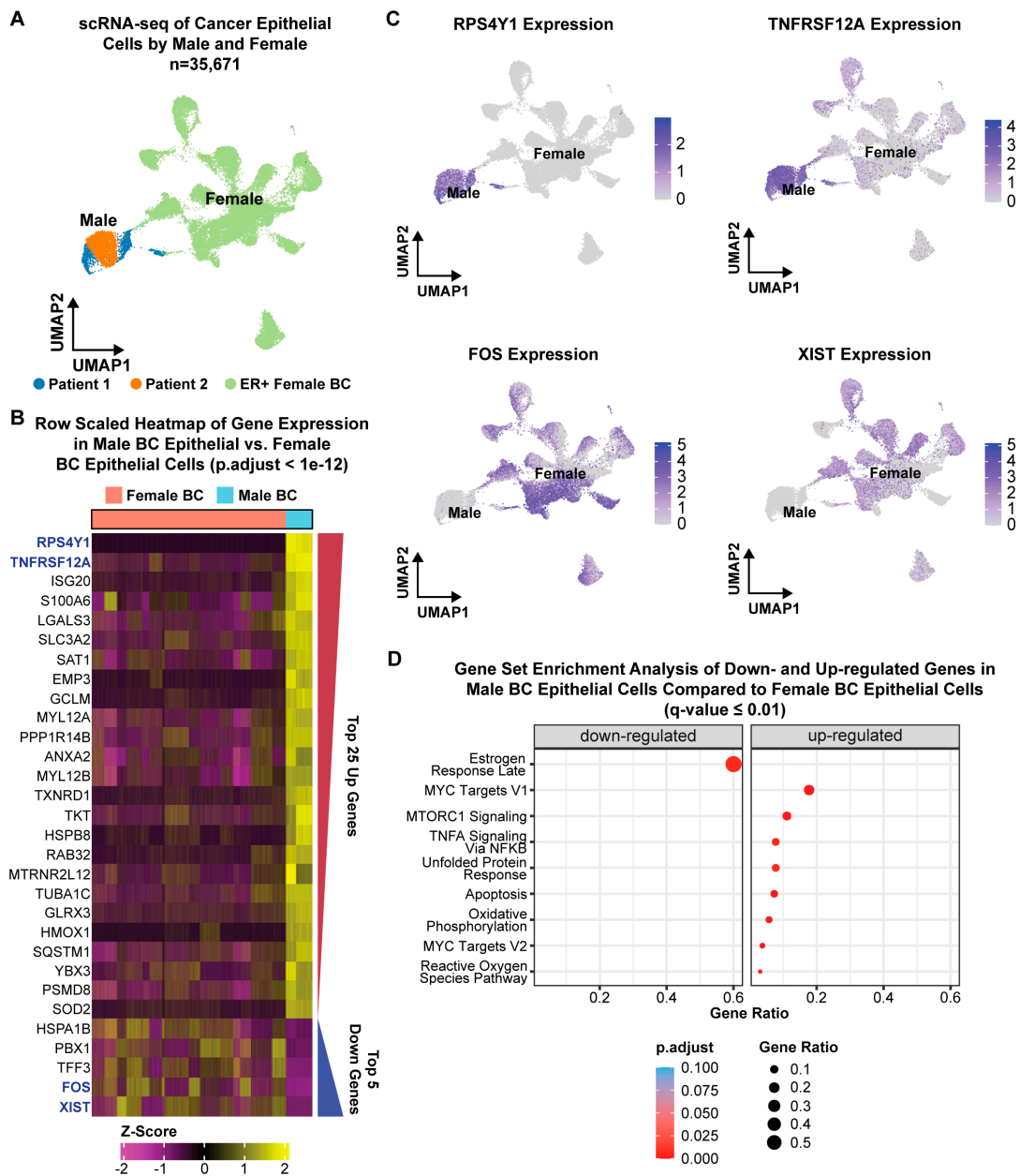


Figure 2. scRNA-seq Analysis Comparing Male and Female Breast Cancer. (A) scRNA-seq UMAP plot of epithelial cells colored by male patients from this study and female patients from Pal et al. [43]. (B) Row-scaled heatmap of gene expression in male cancer epithelial cells versus female cancer epithelial cells. The heatmap shows the top 25 upregulated genes and the top 5 downregulated genes with p -value adjusted $< 1 \times 10^{-12}$. Red/blue wedges to the right of the heatmap correspond to the p -value adjusted trends among the upregulated and downregulated groups. More significant upregulated genes are at the top of the heatmap while the most significant downregulated genes are at the bottom of the heatmap. (C) UMAP plots of scRNA-seq male and female cancer epithelial cells colored by normalized expression of the two most significantly upregulated genes in male breast cancer (*RPS4Y1* and *TNFRSF12A*) and the two most significantly downregulated genes in male breast cancer (*FOS* and *XIST*) compared to female breast cancer. (D) Hallmark gene set enrichment analysis of downregulated and upregulated genes in male breast cancer tumors compared to female breast cancer tumors ($q\text{-value} \leq 0.01$).

Conversely, the downregulated genes were enriched in the 'Estrogen Response Late' pathway, suggesting that although the male epithelial cells are ER+, the estrogen driven gene expression pathways may be muted. Together, these data reveal sex-specific gene expression differences specifically within the malignant cell types of MBC and FBC.

2.3. Systematic Annotation of Cancer-Specific Enhancer Elements in MBC

Chromatin accessibility is a prerequisite for gene transcription and for the activity of the regulatory elements that regulate transcription. Moreover, chromatin accessibility is cell-type specific, usually constrained to the gene expression patterns of a particular cell lineage. Thus, it is possible to deconvolute the cis-regulatory elements that drive cell-type-specific gene expression within MBC tumors using scATAC-seq. First, we called statistically significant chromatin accessibility peaks across all cells and found that the majority of peaks were located in intronic regions (46.2%) or distal intergenic regions (31.3%) (Figure 3). Next, we linked changes in chromatin accessibility to changes in gene expression by performing a peak-to-gene correlation analysis (see Methods; Figure 3) [39]. Briefly, we aggregated the sparse peak counts within groups of similar scATAC-seq cells (~100 cells per group), identified via k-nearest neighbors, to generate more informative metacell observations for each peak in the analysis. Then we computed the correlation between the accessibility of every peak and the expression of every gene across scATAC-seq cells imputed after the Seurat v4.0.15 label transfer procedure [40]. Overall, the peak-to-gene linkage analysis identified 11,719 unique distal peaks participating in 22,869 distal peak-to-gene links in cis across all cell types (Pearson correlation ≥ 0.45 and FDR $< 1 \times 10^{-12}$). We further categorized these peak-to-gene links into five k-means clusters before visualizing them in heatmap form, where we observed highly consistent patterns between gene expression and linked peak accessibility (Figure 3B). Furthermore, the k-means clustering of the peak-to-gene links revealed cell-type-specific enhancer-gene pairs that were characteristic of each cell type; macrophages (cluster 1), epithelial-tumor cells (clusters 2 and 3), fibroblasts (cluster 4), and endothelial cells (cluster 5) (Figure 3B). Herein, we refer to these linked distal peaks as putative enhancers of their paired target genes (Supplementary Table S2).

To differentiate between cancer-specific enhancers and lineage-specific enhancers, we extracted the genomic coordinates for each peak that was enriched in the cancer-epithelial cell-specific k-means groups (clusters 2 and 3; Figure 3B) and overlapped these with previously annotated enhancer elements in normal human mammary epithelial cells (HMECs) [49]. We found a total of 5141 cancer-specific enhancers (participating in 11,551 peak-to-gene links) that were not present in normal mammary epithelial cells and specifically active in the cancer-epithelial cells of MBC (Figure 3D, Supplementary Table S2). Interestingly, the cancer-specific enhancers tend to link to more genes on average (~2.2 genes per enhancer) compared to the non-cancer enhancers found in all other cell types (~1.7 genes per enhancer) (Figure 3E). This suggests that the cancer-specific enhancers may have a higher regulatory load, being associated to the expression of more genes, as compared to lineage-specific enhancers. We also note that transcription factor motif enrichment analysis of the cancer-specific enhancers revealed *FOXA1*, *MESP1/2*, and *TEAP2C* as the top three transcription factors enriched at these enhancers based on adjusted *p*-value (Supplementary Table S3). These data point to a significant rewiring of enhancer elements in cancer-epithelial cells compared to normal epithelial cells, that can potentially sustain oncogenic transcription in MBC.

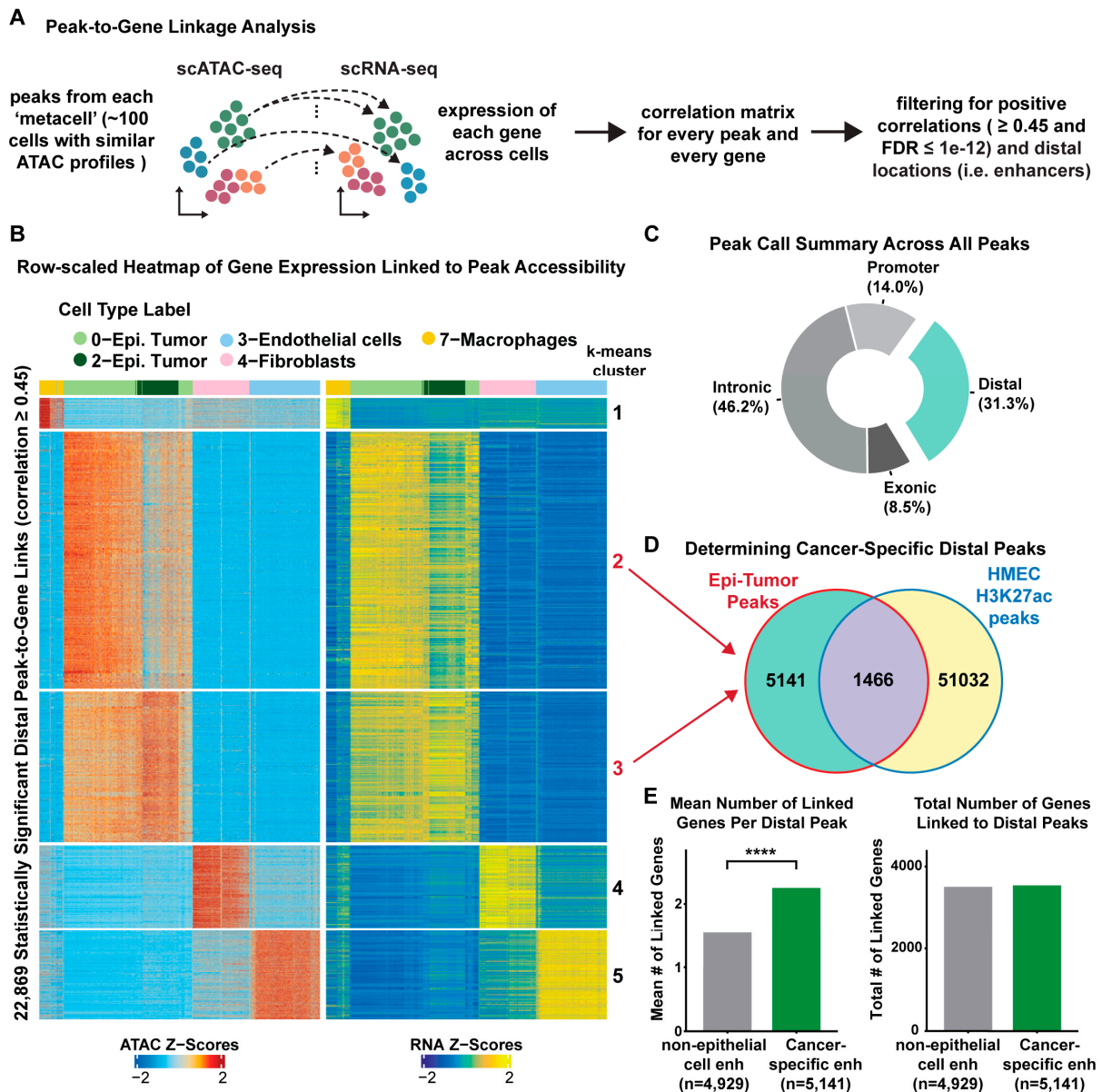


Figure 3. Identification of cancer-specific enhancers in male breast cancer. (A) Workflow for the enhancer peak-to-gene identification strategy. (B) Row-scaled heatmap of statistically significant distal peak-to-gene links ($FDR < 1 \times 10^{-12}$). Rows represent the accessibility of a distal peak and the expression of its linked gene. Columns represent cell types. Cancer-epithelial peak-to-gene links are clustered by k-means clustering and corresponding cluster numbers are denoted in red at the right side of the heatmap. (C) Peak call summary pie chart showing the location of all peaks. (D) Venn diagram showing the number of cancer-specific distal peaks (i.e., 5141 peaks in clusters 2 and 3 from (B)) that do not overlap with regulatory elements found in normal human mammary epithelial cells (HMEC). (E) *Left*: Bar chart depicting the mean number of linked genes per distal peak in cancer-specific distal peaks (green, $n = 5141$) vs. all non-epithelial peaks (grey, $n = 4929$) cells. **** denote p -value $< 2.22 \times 10^{-16}$ (Wilcoxon rank-sum test). *Right*: Bar chart depicting the total number of genes linked to cancer-specific distal peaks (3586 total genes) vs. non-epithelial distal peaks (3551 total genes).

2.4. Tumor Epithelial Cells Acquire Enhancer Elements That Drive Genes Involved in Cancer Progression

The peak-to-gene analysis revealed a striking amount of cancer-specific enhancer-gene pairs that could potentially serve as biomarkers or even tractable pathways in MBC. Therefore, we wanted to identify the cancer-specific enhancers that were linked to MBC genes that are upregulated in comparison to FBC. In total, we found 61 genes that are upregulated in MBC compared to FBC, whose enhancers are specifically active in the cancer-epithelial cell fraction of MBC tumors (Figure 4A,B). One of the highest expressed genes is *ANXA2*, a gene involved in tumor heterogeneity and cancer progression [50]. It has been reported that *ANXA2* is more expressed in African American triple negative breast cancer (TNBC) patients compared to Caucasian TNBC patients, and high expression of *ANXA2* is associated with worse survival [51]. The enhancer linked to *ANXA2* is significantly enriched in the cancer-epithelial cells but not in the non-malignant cell types ($\log_2\text{FC} = 2.3$ and $\text{FDR} < 0.001$), resulting in significantly higher expression of *ANXA2* as measured by the scRNA-seq (Wilcoxon rank sum tests, $p\text{-value} < 2.22 \times 10^{-16}$) (Figure 4C). Of note, this enhancer is not normally active in human mammary epithelial cells (HMEC) and, perhaps more interestingly, is also not annotated in the ENCODE Consortium's curated registry of cis-regulatory elements compiled across hundreds of different cell types [49] (Figure 4C). Furthermore, motif analysis revealed YY1 as the most significantly enriched transcription factor at this enhancer and MAFF as the highest enriched transcription factor at the promoter of *ANXA2* (Figure 4D). YY1 is an important mediator of enhancer-promoter interactions [52], and the expression of YY1 and MAFF in the cancer-epithelial cells was confirmed using the matching scRNA-seq dataset (Figure 4D). Taken together, this previously un-annotated cancer-specific enhancer represents a novel regulator of *ANXA2* in MBC.

Another salient example of a cancer-specific enhancer linked to a male breast cancer upregulated gene is *PRDX4*. We selected *PRDX4* for this example because this gene is involved in breast cancer metastasis [53,54]. We found that *PRDX4* is more highly expressed in MBC compared to FBC ($\log_2\text{FC} > 1.0$), and its expression in cancer-epithelial cells is significantly higher than in the non-malignant cell types (Figure 4E). In this case, there are three enhancer peak-to-gene links to the *PRDX4* promoter (Figure 4E). Among these three peak-to-gene links, enhancer 2 showed the highest enrichment in the cancer epithelial cells compared to the non-malignant cells ($\log_2\text{FC} = 3.5$ and $\text{FDR} < 1 \times 10^{-5}$) and was also not present in normal mammary epithelial cells (Figure 4E). Motif search revealed that FOXP1, TFAP2A, and RUNX1 were the most significantly enriched transcription factors in enhancers 1–3, respectively, and ZKSCAN1 was the most enriched transcription factor at the promoter of *PRDX4* (Figure 4F). Finally, we confirmed the expression of these transcription factors specifically within cancer epithelial cells using the scRNA-seq data, giving confidence to our results.

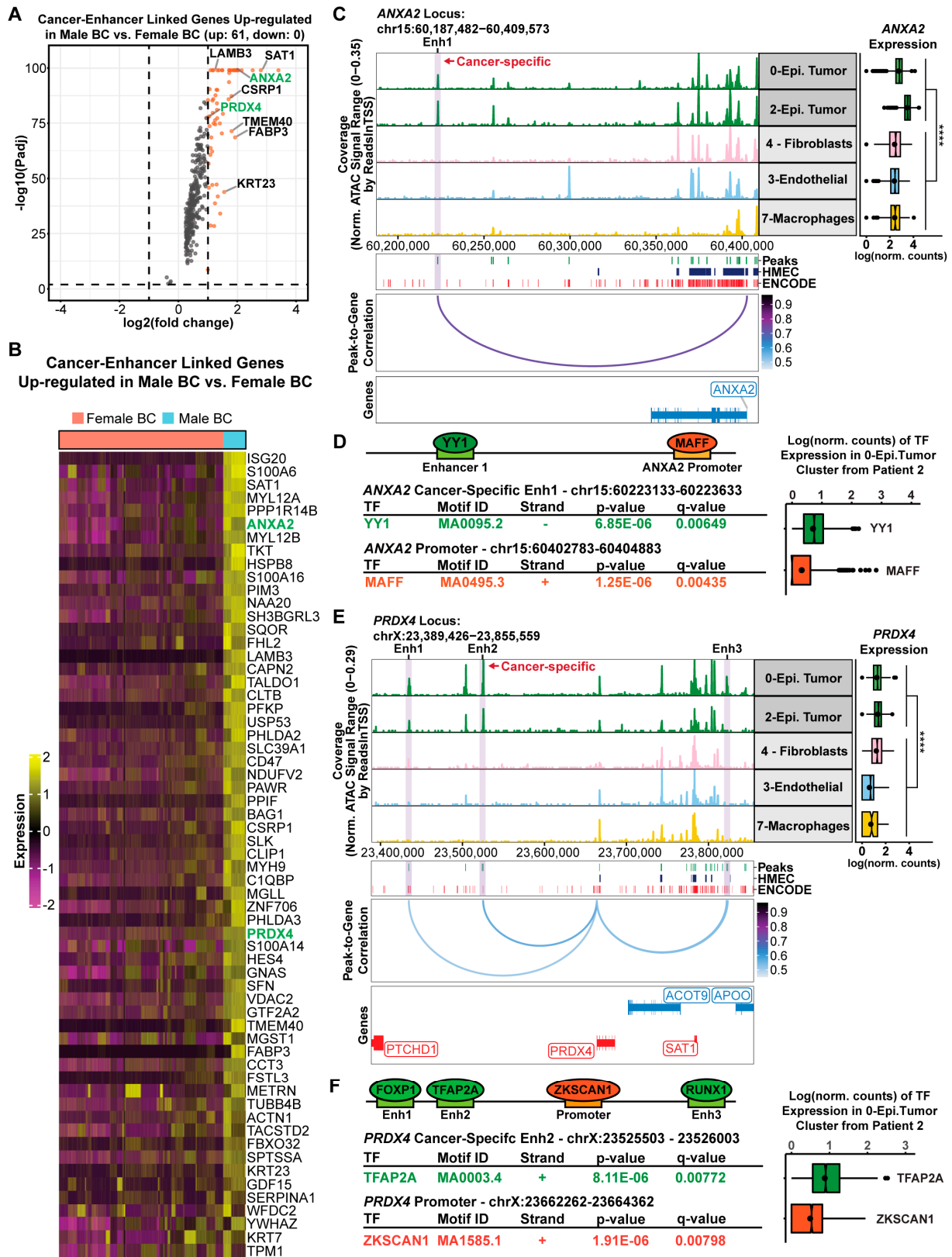


Figure 4. Cancer-specific enhancers drive the expression of ANXA2 and PRDX4 in male breast cancer. (A) Volcano plot of cancer-enhancers linked to genes that are upregulated in male breast cancer cells versus female breast cancer cells. (B) Row-scaled heatmap of all 61 upregulated cancer-enhancer

linked genes in male breast cancer versus female breast cancer. More significant upregulated genes are at the top of the heatmap. (C) Browser track view of chromatin accessibility at the ANXA2 locus across epithelial cancer cells (Epi. Tumor) and all other cell-type clusters. The cancer-specific enhancer 1 (Enh1) peak that is linked to ANXA2 gene expression is highlighted in light purple. The peak track below the browser track denotes all scATAC-seq peaks from this study (Peaks), regulatory elements found in human mammary epithelial cells (HMEC), and all ENCODE cis-regulatory elements. The peak-to-gene correlation loops show the correlation between ANXA2 and Enh1. Gene expression of ANXA2 in matched scRNA-seq cells is depicted to the right of the browser track. **** denote p -value $< 2.22 \times 10^{-16}$ (Wilcoxon rank-sum test). (D) Find individual motif occurrences (FIMO) predictions within the ANXA2 cancer-specific Enh1 and the ANXA2 promoter showing the predicted transcription factors with the highest expression in the 0-Epi.Tumor cluster from patient 2. The gene expression of each TF specifically within the cancer epithelial cells is shown to the right of the table. (E) Browser track view of chromatin accessibility at the PRDX4 locus across epithelial cancer cells (Epi. Tumor) and all other cell-type clusters. Predicted enhancers of PRDX4 are highlighted in light purple. The cancer-specific peak, Enh2, is denoted in red. The peak track below the browser track denotes all scATAC-seq peaks from this study (Peaks), regulatory elements found in human mammary epithelial cells (HMEC), and all ENCODE cis-regulatory elements. The peak-to-gene correlation loops show the correlation between PRDX4, and the peaks linked to this gene. Gene expression of PRDX4 in matched scRNA-seq cells is depicted to the right of the browser track. **** denote p -value $< 2.22 \times 10^{-16}$ (Wilcoxon rank-sum test). (F) Find individual motif occurrences (FIMO) predictions within the PRDX4 cancer-specific Enh2 and the PRDX4 promoter showing the predicted transcription factors with the highest expression in the 0-Epi.Tumor cluster from patient 2. The gene expression of each TF specifically within the cancer epithelial cells is shown to the right of the table.

As we were looking through the 61 MBC genes that are linked to cancer-specific enhancers, we noticed a few instances of clusters of neighboring enhancers that had abnormally high levels of chromatin accessibility in the cancer-epithelial cell types. These clusters of enhancers were highly active and highly interconnected, reminiscent of super-enhancer function [55,56]. For example, *LAMB3* is very highly expressed in MBC epithelial tumor cells, with more than five peak-to-gene links associated with its promoter (Supplementary Figure S3A). The cluster of enhancer peaks near chr1:209,444,222-09,546,602 were significantly correlated to each other and individually linked to the promoter of *LAMB3*, suggesting that this may be a super-enhancer (Supplementary Figure S3A). *LAMB3* has been previously related with focal adhesion for cell migration in breast cancer [57], and its higher expression is associated with worse overall survival (OS) and disease-free survival (DFS) in pancreatic ductal adenocarcinoma (PDAC) [58]. Therefore, we posit high expression of *LAMB3* is unfavorable for MBC patients and its elevated expression is sustained by the super-enhancer. Interestingly, we found a second cluster of cancer-cell-specific enhancers that are linked with the expression of *CD47* and overlap with a previously annotated super-enhancer in breast cancer [59] (Supplementary Figure S3B). As seen with the previous example (*LAMB3*), the cluster of enhancer peaks near chr3:107,999,565-108,001,642, were highly interconnected and linked to the promoter of *CD47* leading to high levels of *CD47* gene expression (Supplementary Figure S3B). *CD47* is a key mediator of immune evasion and epithelial to mesenchymal transition in breast cancer and its high expression is related to worse disease-free survival [60]. Although we did not intentionally set out to find super-enhancers, these examples highlight the potential for finding ‘super-enhancers’ at single-cell resolution. These data begin to explain how differentially expressed genes in MBC are regulated by cancer-specific enhancers, highlighting their potential as novel targets for therapeutic intervention.

3. Discussion

The results described herein exemplify the utility of our matched scRNA-seq and scATAC-seq dataset for uncovering clinically relevant mechanisms of gene expression in MBC. We recognize that our study is limited to two ER+ male breast cancer patients, but it is important to note that MBC represents only 1% of all breast cancer diagnoses, making it difficult to procure these rare tumor types. Moreover, we had the additional requirement of collecting live tumor specimens, on the day of surgical resection, to ensure high quality single-cell datasets. Thus, our study represents the first matched scRNA-seq and scATAC-seq dataset of MBC and represents a foundation for determining the regulatory logic of MBC. This is important because the 5-year mortality of MBC is higher than that of FBC [4] and not enough is known about MBC at the molecular level, especially at single-cell resolution.

While the anatomical listing of cell types within MBC tumors was not the initial goal of this project, our high-quality scRNA-seq dataset enabled the detection of various cell types, including macrophages, fibroblasts, endothelial cells, and epithelial cells. The malignant cell types were detected by inferring copy-number variation from the scRNA-seq data, and notably we found that the non-malignant cell types showed similar gene expression profiles across patients (suggesting a similar biological state) while the cancer-epithelial cells showed more patient-specific transcriptomes, pointing to the heterogeneity of MBC. Comparison of the cancer-epithelial cell transcriptomes to the female cancer-epithelial cell transcriptomes revealed distinctive gene expression programs as evidenced by the discrete clustering of the male versus female epithelial cells (Figure 2). The top five upregulated genes were *RPS4Y1*, *TNFRSF12A*, *ISG20*, *S100A6*, and *LGALS3*, whereas the top five downregulated genes were *XIST*, *FOS*, *TFE3*, *PBX1*, and *HSPA1B*. *RPS4Y1* was not expressed in female epithelial cells since it was located on the Y-chromosome, confirming the soundness of differential gene expression analysis. Interestingly, the other upregulated genes are associated with features of cell proliferation (*TNFRSF12A* in human hepatocellular carcinoma [61]), tumor progression (*ISG20* in clear cell renal cell carcinoma [62]), and cell survival (*LGALS3* in breast cancer [63]). Collectively, the pathway analysis of all upregulated genes shows enrichment in hallmark pathways such as ‘MYC targets’ and ‘mTORC1 signaling’ (Figure 2). Indeed, the upregulation of the *mTOR* pathway in male breast cancer has been shown before using bulk microarray studies [23], suggesting that MBC patients may benefit from *mTOR* inhibitors. Conversely, the downregulated genes were enriched in the ‘Estrogen Response Late’ hallmark pathway, suggesting that although the male epithelial cells are ER+, the estrogen-driven gene expression pathways may be muted. Similar to FBC, endocrine therapy targeting ER is the mainstay of systemic therapy for ER+ MBC, and these results indicate that MBC may be inherently less estrogen responsive. No data currently exist prospectively determining the efficacy of endocrine therapy in MBC patients. The ongoing ETHAN prospective, randomized trial (<https://clinicaltrials.gov/ct2/show/NCT05501704> (accessed on 17 April 2023)), will compare endocrine therapy regimens in male breast cancer patients and this will be important to establish the role of endocrine therapy for male patients. Additionally, this trial will provide samples for translational correlates of how MBC respond to endocrine therapy, and to test how expression of genes identified in our study correlate with response.

The chromatin accessibility profiles enabled the systematic detection of regulatory elements that drive the distinctive gene expression programs within MBC. We demonstrated that there is widespread rearrangement of the enhancer landscape in MBC and that cancer-epithelial cells acquire de novo enhancer elements that drive the expression of male-specific genes. Starting from a broad survey of cis-regulatory elements across all cell-types, we were able to identify cancer-cell-specific enhancer-to-gene links that are not typically active in normal cell types (Figure 3). Examples such as *ANXA2* and *PRDX4* show significant enrichment in the chromatin accessibility of the cancer-specific enhancers compared to the non-malignant cell types, resulting in marked increases in gene expression (Figure 4). Previous studies in FBC have shown that increased *ANXA2* gene expression is associated

with drug resistance and tumor recurrence [64]. Since *ANXA2* expression is even higher in MBC, our enhancer–promoter interaction model may provide a way to understand the mechanism by which MBC becomes resistant to standard therapy. Similarly, *PRDX4* is overexpressed in several cancers such as gastric cancer [65], lung cancer [66], and breast cancer [54]. In addition, elevated *PRDX4* expression is associated with worse overall and disease-free survival in FBC patients [54]. Thus, the heightened expression of *PRDX4* in males versus females may underlie observed worse outcome for male patients.

Perhaps more interestingly, we saw evidence for super-enhancer activity at single-cell resolution. Super-enhancers are defined as large clusters of neighboring enhancers that have an unusually high occupancy of interacting factors and are thought to act synergistically with each other to promote the expression of their target genes [55]. Super-enhancers have gained much attention because they are known to regulate key cell identity genes, and in cancer, are known to drive oncogene expression [56,67]. Within our list of 61 upregulated male versus female breast cancer genes linked to cancer-specific enhancers, we identified at least two super-enhancers; one linked to *LAMB3* and the other to *CD47* (Supplemental Figure S3). Notably, the super-enhancer linked to *CD47* has been annotated before in breast cancer [59] and we show evidence of the increased activity of the constituent enhancers specifically within the cancer-epithelial cell types (Supplemental Figure S3). *CD47* is a transmembrane protein that belongs to the immunoglobulin superfamily that enables binding adhesion to the extracellular matrix. The encoded protein binds the ligand thrombospondin and plays a role in membrane signal transduction [60,68]. In breast cancer, *CD47* is associated with epithelial-mesenchymal transition and poor DFS [60,68]. Therefore, the high expression of *CD47* in MBC may suggest a possibility for targeting *CD47* as a therapeutic strategy to overcome immune evasion [68]. Although we may be underpowered to comprehensively annotate super-enhancers using single-cell data, the detection of super-enhancers at single-cell resolution is an interesting concept and inspires further investigation. Taken together, this study annotates clinically relevant regulatory networks in male breast tumors at single-cell resolution, providing a useful resource that expands our current understanding of the gene expression programs that underlie the biology of MBC.

4. Materials and Methods

4.1. Human Patient Tumor Dissociation

Fresh, never frozen or fixed, tumors were collected by UNC's Tissue Procurement Facility and immediately transported to the lab on ice in DMEM/F12 media + 1% penicillin/streptomycin after surgical resection or core biopsy. Tumor specimens were dissociated and prepped as described in our previous work [37]. Briefly, tumor specimens were minced with two sterile razor blades and incubated overnight on a 37 °C stir plate at 180 rpm in the following digestion media: DMEM/F12, 5% FBS, 15 mM HEPES, 1× Glutamax, 1× Gentle Collagenase/Hyaluronidase (Stem Cell Technologies, Vancouver, BC, Canada, 07919), 1% penicillin/streptomycin, and 0.48 µg/mL Hydrocortisone (Stem Cell Technologies, 74144). After overnight digestion, tumor cells were pelleted at 1200 rpm for 5 min at room temperature and washed two times with cold PBS + 2% FBS and 10 mM HEPES (PBS-HF). Removal of red blood cells was done by treating cell pellets with cold ammonium chloride solution (Stem Cell Technologies, 07850) diluted with 1 part PBS-HF. To ensure complete dissociation of tumor cells and removal of free DNA in the cell suspension, cells were treated with 0.05% Trypsin-EDTA and 200 µL 1 mg/mL DNase I for one minute, followed by trypsin inactivation with PBS-HF and centrifugation. Cells were then washed with PBS-HF, filtered through a 100µm cell strainer and a final centrifugation was done. Cell pellets were resuspended in DMEM/F12 + 5% FBS using a volume that was dependent on the size of the final cell pellet, then filtered through a 40µm cell strainer. We measured cell viabilities for two patient samples with the Countess II Automated Cell Counter (Thermo Fisher (Waltham, MA, USA) AMQAX1000), cell viabilities were 90% for the Patient 1 sample and 88% for the Patient 2 sample.

4.2. Single-Cell Sequencing Library Preparation

Immediately following dissociation, cells were diluted to 1200 cells/ μ L in preparation for scRNA-seq and scATAC-seq library generation as described in our previous work [37]. For scRNA-seq libraries, 10,000 cells were used for library preparation using the 10 \times Genomics Single Cell 3' kits: Single Cell 3' GEM, Library & Gel Bead Kit v3 (PN-1000075), Chromium Chip B Single Cell Kit (PN-10000153), and Chromium i7 Multiplex Kit (PN-120262) following the manufacturer's protocol.

For scATAC-seq, nuclei isolation of 500,000 cells was carried out following the Nuclei Isolation for Single Cell ATAC Sequencing protocol from 10 \times Genomics using a four-minute lysis time. Next, 10,000 nuclei were used for library preparation using the 10 \times Genomics Single Cell ATAC Kits: Chromium Single Cell ATAC Library & Gel Bead Kit v1 (PN-1000110), Chromium Chip E Single Cell ATAC Kit (PN-1000082), and Chromium i7 Multiplex Kit N, Set A (PN-1000084) following the manufacturer's protocol. scRNA-seq and scATAC-seq libraries were sequenced using 10 \times Genomics' suggested sequencing parameters on an Illumina NextSeq 500 machine by UNC's Translational Genomics Lab.

4.3. Single-Cell RNA-seq Quantification and Quality Control (QC)

Cell Ranger from 10 \times Genomics was used to generate raw and filtered feature barcode matrices for each patient sample. A Seurat object was built from the filtered feature barcode matrix for each patient sample by using the Seurat R package [41,69]. Quality control (QC) and doublet removal were performed separately for each patient dataset to select for high quality cells. First, outlier cells were defined in each of the following metrics: number of UMI counts (<5000), number of genes expressed (<2000) and percent mitochondrial read count (>25%). Outlier cells according to these criteria were removed before doublet detection. Then, doublet removal was performed by DoubletFinder [70]. After QC and doublet removal for each patient dataset, we applied Seurat's *merge()* function [40] to combine the individual patient datasets, forming the male breast cancer cohort presented in this study.

4.4. Single-Cell RNA-seq Normalization, Feature Selection and Clustering

Seurat's *NormalizeData()* function [40], with the normalization method set to "LogNormalize", was used to normalize gene expression matrices. Seurat's *FindVariableFeatures()* function [40], with the selection method set to "vst" and the number of top variable features set to 2000, was used to perform feature selection. Seurat's *ScaleData()* function [40] was used to scale the expression values for the top 2000 variably expressed genes in the dataset before carrying out principal component analysis (PCA). We chose to regress out the percentage of mitochondrial genes when using Seurat's *ScaleData()* function [40]. The top 2000 most variably expressed genes were summarized by PCA into 50 principal components (PCs). Cells were then visualized in a UMAP embedding with Seurat's *RunUMAP()* function [40] using 30 PCs. Next, the shared nearest-neighbor graph was constructed by Seurat's *FindNeighbors()* function [40] using 30 PCs and Louvain clustering was performed with Seurat's *FindClusters()* function [40] with a resolution of 0.8. scRNA-seq UMAP embeddings were plotted in R [69] using ggplot2 [71].

4.5. Inference of Copy Number Variation (CNV) from Single-Cell RNA-Seq

Within each patient sample, estimated copy number events for each cell cluster were derived using the R package inferCNV [42]. Immune cell and endothelial cell clusters were used as a normal background for inferCNV. The remaining cell clusters were specified in the inferCNV annotations file to infer CNVs at the level of these clusters. The standard inferCNV algorithm was invoked with *infercnv::run()* with the following parameters of (cutoff: 0.1, scale_data: FALSE, HMM: FALSE, and denoise: TRUE). Epithelial cells were classified into epithelial tumor, epithelial unassigned, epithelial normal after plotting scatter plots with CNV values and correlation with top 5% cells with high CNV values [72,73].

Among nine clusters, two clusters were assigned as epithelial unassigned, and four clusters were assigned as epithelial tumor (Supplementary Figure S2B).

4.6. Single-Cell RNA-seq Cell Type Annotation

The cell types were annotated with the R package SingleR [74] based on reference transcriptomic datasets of pure cell types and gene signature enrichment obtained from Seurat's *AddModuleScore()* function [40]. The normalized expression values built from human bulk RNA-seq generated and supplied by Blueprint and ENCODE were used as a reference dataset for SingleR [74], which is available via R package celltex [74]. We built a MBC patient cohort dataset after merging the raw count matrices of both patients into a Seurat object by using Seurat's *merge()* function [40], we applied Harmony batch correction [38], and finally, cells were clustered based on the Louvain algorithm (described above). The resulting cell-type clusters in our merged dataset were assigned a cell type labeled based on the majority cell type in each cluster.

Among nine clusters in the full cohort of two MBC patients, six clusters were assigned as epithelial clusters since the most frequent cells of these clusters were epithelial cells. The other three clusters were a fibroblast cluster, an endothelial cluster, and a macrophage cluster. The mean number of RNA counts across all nine clusters was larger than 5000 (Supplementary Figure S1C), since outlier cells with UMI counts < 5000 were removed at the QC step. Similarly, the mean number of features across for nine clusters was greater than 2000 (Supplementary Figure S1D), since outlier cells with features < 2000 were also removed at the QC step.

4.7. Differentially Expressed Genes in Single-Cell RNA-Seq between Female BC and Male BC

ScRNA-seq data for female ER+ breast cancer data were downloaded from GSE161529 [43]. The differentially expressed genes in single cells between 16 female ER+ breast cancer patients and two ER+ male breast cancer patients were identified by Seurat's *FindMarkers()* function [40] with parameters (logfc.threshold; 0.25, min.pct: 0.5, min.diff.pct: 0.25, and max.cells.per.ident: 500). Then, markers were selected by an additional filter of $\text{Padj} < 0.01$. The heatmap of Figure 2B includes top 25 upregulated genes in male BC and top 5 downregulated genes with additional criterion of $\text{Padj} < 1 \times 10^{-12}$ to select smaller number of genes for visualization. All of the differentially regulated genes are tabulated in Supplemental Table S1. The enrichment of cancer hallmark gene sets [45] were identified by hypergeometric tests with clusterProfiler *enricher()* [75] with $q\text{-value} \leq 0.01$.

4.8. Single-Cell ATAC-seq Quality Control (QC)

Cell Ranger ATAC from 10× Genomics was used to demultiplex raw base call files into FASTQ files and generate a filtered peak-barcode matrix containing detected cellular barcodes and a fragments file as in the BED format for each patient sample. These fragment lists were read into the R package ArchR [39] to perform quality control and doublet removal. To enrich for cellular barcodes, a threshold for $\log_{10}(\text{TSS enrichment} + 1)$ was set manually to 0.9 for both scATAC-seq samples while a sample-specific threshold of $\log_{10}(\text{number of unique fragments})$ was estimated using a Gaussian Mixture Model (GMM) for each scATAC-seq sample, as implemented in the R package mclust [76]. Barcodes below these thresholds in any of these metrics were excluded before doublet detection step. ArchR's *addDoubletScores()* function [39], with the *knnMethod* parameter of "UMAP", was used to estimate doublet enrichment scores, and ArchR's *filterDoublets()* function [39], with the *filterRatio* parameter of 1.0, was used to filter out cellular barcodes as doublets.

4.9. Single-Cell ATAC-seq Quantification, Feature Selection and Integration with Single-Cell RNA-seq

An initial tile matrix of 500 bp genomic tiles across all cells was generated by the ArchR package [39]. We used the iterative latent semantic indexing [35,36,77] (LSI) procedure implemented in the ArchR R package to reduce dimensions of the genomic tile matrix using two iterations with 25,000 variable features. 30 LSI dimensions were used to create a UMAP embedding with ArchR's *addUMAP()* [39] with the reduced dimension object obtained by Harmony batch correction [38,39]. By applying a *corCutOff* parameter of 0.75 to ArchR's *addIterativeLSI()*, we excluded LSI dimensions that have a correlation to sequencing depth greater than 0.75. scATAC-seq UMAP embeddings were plotted in R [69] using *ggplot2* [71].

ArchR's *addGeneScoreMatrix()* function was used to estimate gene activity scores by considering accessibility within the entire gene body and the activity of putative distal regulatory elements [39]. Seurat's CCA implementation [41] was executed via ArchR's *addGeneIntegrationMatrix()* function [39] to integrate scATAC cells with a scRNA cells by assigning each of the scATAC-seq cells a cell-type cluster identity from the matching scRNA-seq data, an associated label prediction score, and an imputed transcriptome. Cells with label prediction scores less than or equal to 0.5 were excluded before obtaining marker features for each cell-type cluster and calling peaks from pseudo-bulk replicates. For each inferred cell-type cluster, pseudo-bulk replicates were generated using the R package ArchR [39] and pseudo-bulk peak calling was performed using MACS2 [78,79]. Peak calls from each inferred cell-type cluster were merged into a universal peak set using ArchR's default iterative overlap procedure. ArchR's *plotBrowserTrack()* function [39] was used to plot genomic browser tracks visualizing the chromatin accessibility across cell-type clusters, the peak locations of pseudo-bulk ATAC-seq peaks, and peak-to-gene correlations.

We applied an additional filtering step to select highly confident epithelial tumor clusters for the downstream analyses including peak-to-gene correlation analysis, identifying cancer-specific distal peaks, and predicting transcription factor occupancy at select putative enhancer regions. Two clusters assigned as epithelial unassigned (i.e., 1-Epi. Unassigned and 8-Epi. Unassigned) were removed, since these clusters could be a mixture of tumor cells and normal cells. Note that the number of cells in the first epithelial unassigned cluster (i.e., 1-Epi. Unassigned) was around 500 cells and the number of cells in the second epithelial unassigned cluster (i.e., 8-Epi. Unassigned) was less than 500 cells. We did not use "the number of cells < 500" as a filtering criterion, but these two epithelial unassigned clusters might have lower quality in peak calling. One cluster assigned as epithelial tumor (i.e., 6-Epi. Tumor) was removed, since the mean number of ATAC fragments was less than 5000 (Supplementary Figure S1G). Note that this cluster had the lowest number of cells (Supplementary Figure S1F).

4.10. Peak-to-Gene Correlation Analysis

ArchR's *addPeak2GeneLinks()* function [39], with *reducedDims* set to "IterativeLSI" and *dimsToUse* set to "1:30", was used to carry out the peak-to-gene correlation analysis to identify putative regulatory elements by correlating peak accessibility with imputed gene expression in scATAC-seq cells [37]. To circumvent the sparsity of scATAC-seq data, low-overlapping aggregates of scATAC-seq cells were generated via a k-nearest neighbor procedure in the LSI space to ensure robust peak-to-gene associations and reduce noise.

Peak-to-gene links were selected by distal peak location and selection criteria of correlation ≥ 0.45 and FDR $\leq 1 \times 10^{-12}$ with the control of Benjamini–Hochberg (BH) method [80]. The number of remaining distal peaks was 11,719, participating in a total of 22,869 peak-to-gene links. The distal peak-to-gene links were clustered using k-means before being visualized in a heatmap using ArchR's *plotPeak2GeneHeatmap()* function [55].

The mean number of linked genes per distal peak and the total number of genes linked with distal peaks (Figure 3E) was computed from a peak-to-gene metadata table having peak names defined by genomic coordinates, and corresponding gene names. The

distribution of the number of linked genes per distal peak was used to compute p -value with Wilcoxon test to show that the numbers of linked genes per peak in epithelial cells and non-epithelial genes are statistically significant.

4.11. Identifying Cancer-Specific Distal Peaks

To identify putative cancer-specific distal peaks as demonstrated in Figure 3A,D, we used a genomic internal overlap analysis as previously described [37]. The genomic coordinates of the distal peaks participating in the cancer-enriched peak-to-gene links were overlapped with a set of normal active enhancer marker H3K27ac peaks in human mammary epithelial cells (HMECs) obtained from ENCODE [49]. To find any overlaps including 1 bp overlap between the cancer-enriched peaks and the normal peaks in HMECs, we used the function *findOverlapsOfPeaks()* from the ChIPpeakAnno R package [81] with the *minoverlap* parameter of 1. The cancer-specific peaks were defined as cancer-enriched peaks that did not overlap with any of the normal peaks in HMECs.

The cancer-specific distal peaks linked with upregulated genes in MBC compared with FBC were determined by upregulation criteria ($\log_2FC > 1.0$ and adjusted p -value < 0.01). Total of 61 cancer-enhancer link genes were selected (Figure 4A). The gene expression of these 61 genes in FBC and MBC cells were visualized in heatmap form (Figure 4B).

The normal peaks in HMECs and the full list of the candidate cis-Regulatory Elements (cCREs) [49] derived from ENCODE data in hg38 were used for visualization of peaks around genes of interest in Figure 4C,E. The gene expression values (\log normalized count) of *ANXA2* and *PRDX4* in two groups (epithelial tumor clusters vs. non-epithelial clusters) were compared by the Wilcoxon test.

Motif enrichment analysis was performed between cancer-specific distal enhancer peaks and other distal enhancer peaks with ArchR's *addMotifAnnotation()* function [39] using CisBP (Catalog of Inferred Sequence Binding Preferences) [82] and *peakAnnoEnrichment()* function [39] with the cutOff of "FDR < 0.01 & $\log_2FC > 1.0$ ". The top 10 motifs enriched in cancer-specific distal enhancer peaks included *FOXA1* and *TFAP2C* (Supplementary Table S3).

4.12. Predicting Transcription Factor Occupancy at Select Putative Enhancer Regions

The genomic location of the enhancer of *ANXA2* was chr15:60,223,133-60,223,633 (Enh), whereas the genomic locations of three enhancer of *PRDX4* were chrX:23,434,983-23,435,483 (Enh1), chrX:23,525,503-23,526,003 (Enh2), chrX:23,822,316-23,822,816 (Enh3).

Bedtools [83] *getfasta()* was used to extract the sequences of the select putative enhancers in the malignant fraction of Patient 2, as shown in Figure 4D,F, after accounting for single-nucleotide variants relative to the hg38 reference genome. To include single-nucleotide variants from the malignant fraction in our analysis, we used bcftools [84] *mpileup* followed by bcftools [84] *consensus* with a bam file containing fragments only from cellular barcodes present in in the Patient 2 malignant fraction. Cell Ranger's *bamslide* was used to subset a position-sorted BAM file to make a malignant-specific BAM file for each patient sample. We ran Find Individual Motif Occurrences (FIMO) [85] motif scanning with the putative enhancer sequences as input and with JASPAR2020 CORE [86] containing curated and non-redundant transcription-factor-binding motifs for vertebrates. The statistically significant motifs with a q -value < 0.01 were sorted by the expression values of corresponding transcription factors in the malignant cells of Patient 2. The expression value of a transcription factor was calculated by the sum of \log -normalized counts of the transcription factors across malignant cells in scRNA-seq, which was visualized in R [69] using ggplot2 [71].

Supplementary Materials: The supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms241713053/s1>.

Author Contributions: H.L.F., K.W. and P.M.S. conceived and the study. K.W., A.T. and P.M.S. carried out tumor sample procurement and generated the scRNA-seq and scATAC-seq libraries. H.K. designed and performed the computational analysis with input from H.L.F. and M.J.R. The manuscript was written by H.K., K.W., M.J.R. and H.L.F. with input from all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by grants from the NIH/National Cancer Institute (R01-CA273444-01), the Susan G. Komen Breast Cancer Foundation (CCR19608601), and the Department of Defense CDMRP Breast Cancer Research Program (BC180450) to H.L.F.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Raw data (10× FASTQs) and processed data for single-cell RNA-seq data and single-cell ATAC-seq have been deposited at the National Institutes of Health (NIH) Database of Genotypes and Phenotypes (dbGaP) (<https://www.ncbi.nlm.nih.gov/gap/>) and is available under the accession number phs003006.v1.p1. All original code has been deposited on the Github, and it is publicly available at the Github repository maleBC_single_cells (https://github.com/hyunsoo77/maleBC_single_cells). Any additional information regarding data analysis is available from the lead contact (hfranco@med.unc.edu).

Acknowledgments: We thank the patients and their families. We thank the University of North Carolina (UNC) Tissue Procurement Facility for helping acquire tumor specimens, and the UNC Translational Genomics Lab for providing experimental support and helping us sequence libraries. We thank Michele Hayward at the Office of Genomics Research for helping us navigate the institutional review board (IRB) process and the data submission process.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Miao, H.; Verkooijen, H.M.; Chia, K.-S.; Bouchardy, C.; Pukkala, E.; Larønningen, S.; Mellekjær, L.; Czene, K.; Hartman, M. Incidence and outcome of male breast cancer: An international population-based study. *J. Clin. Oncol.* **2011**, *29*, 4381–4386. [[CrossRef](#)] [[PubMed](#)]
2. American Cancer Society. *Cancer Facts & Figures 2022*; National Home Office: Atlanta, GA, USA, 2022.
3. Greif, J.M.; Pezzi, C.M.; Klimberg, V.S.; Bailey, L.; Zuraek, M. Gender differences in breast cancer: Analysis of 13,000 breast cancers in men from the National Cancer Data Base. *Ann. Surg. Oncol.* **2012**, *19*, 3199–3204. [[CrossRef](#)] [[PubMed](#)]
4. Wang, F.; Shu, X.; Meszoely, I.; Pal, T.; Mayer, I.A.; Yu, Z.; Zheng, W.; Bailey, C.E.; Shu, X.-O. Overall Mortality After Diagnosis of Breast Cancer in Men vs Women. *JAMA Oncol.* **2019**, *5*, 1589–1596. [[CrossRef](#)] [[PubMed](#)]
5. Giordano, S.H. Breast Cancer in Men. *N. Engl. J. Med.* **2018**, *379*, 1385–1386. [[CrossRef](#)]
6. Cardoso, F.; Bartlett, J.; Slaets, L.; van Deurzen, C.; van Leeuwen-Stok, E.; Porter, P.; Linderholm, B.; Hedenfalk, I.; Schröder, C.; Martens, J.; et al. Characterization of male breast cancer: Results of the EORTC 10085/TBCRC/BIG/NABCG International Male Breast Cancer Program. *Ann. Oncol.* **2018**, *29*, 405–417. [[CrossRef](#)]
7. Doebar, S.C.; Slaets, L.; Cardoso, F.; Giordano, S.H.; Bartlett, J.M.; Tryfonidis, K.; Dijkstra, N.H.; Schröder, C.P.; van Asperen, C.J.; Linderholm, B.; et al. Male breast cancer precursor lesions: Analysis of the EORTC 10085/TBCRC/BIG/NABCG International Male Breast Cancer Program. *Mod. Pathol.* **2017**, *30*, 509–518. [[CrossRef](#)]
8. Vera-Badillo, F.E.; Templeton, A.J.; De Gouveia, P.; Diaz-Padilla, I.; Bedard, P.L.; Al-Mubarak, M.; Seruga, B.; Tannock, I.F.; Ocana, A.; Amir, E. Androgen receptor expression and outcomes in early breast cancer: A systematic review and meta-analysis. *J. Natl. Cancer Inst.* **2014**, *106*, djt319. [[CrossRef](#)]
9. Piscuoglio, S.; Ng, C.K.; Murray, M.P.; Guerini-Rocco, E.; Martelotto, L.G.; Geyer, F.C.; Bidard, F.-C.; Berman, S.; Fusco, N.; Sakr, R.A.; et al. The Genomic Landscape of Male Breast Cancers. *Clin. Cancer Res.* **2016**, *22*, 4045–4056. [[CrossRef](#)]
10. Johansson, I.; Nilsson, C.; Berglund, P.; Lauss, M.; Ringnér, M.; Olsson, H.; Luts, L.; Sim, E.; Thorstensson, S.; Fjällskog, M.-L.; et al. Gene expression profiling of primary male breast cancers reveals two unique subgroups and identifies N-acetyltransferase-1 (NAT1) as a novel prognostic biomarker. *Breast Cancer Res.* **2012**, *14*, R31. [[CrossRef](#)]
11. Johansson, I.; Ringnér, M.; Hedenfalk, I. The landscape of candidate driver genes differs between male and female breast cancer. *PLoS ONE* **2013**, *8*, e78299. [[CrossRef](#)]
12. Friedman, L.S.; Gayther, S.A.; Kurosaki, T.; Gordon, D.; Noble, B.; Casey, G.; Ponder, B.A.; Anton-Culver, H. Mutation analysis of BRCA1 and BRCA2 in a male breast cancer population. *Am. J. Hum. Genet.* **1997**, *60*, 313–319.

13. Basham, V.M.; Lipscombe, J.M.; Ward, J.M.; Gayther, S.A.; Ponder, B.A.; Easton, D.F.; Pharoah, P.D. BRCA1 and BRCA2 mutations in a population-based study of male breast cancer. *Breast Cancer Res.* **2002**, *4*, R2. [[CrossRef](#)]
14. Sung, H.; DeSantis, C.; Jemal, A. Subtype-Specific Breast Cancer Incidence Rates in Black versus White Men in the United States. *JNCI Cancer Spectr.* **2020**, *4*, pkz091. [[CrossRef](#)] [[PubMed](#)]
15. Brinton, L.A.; Cook, M.B.; McCormack, V.; Johnson, K.C.; Olsson, H.; Casagrande, J.T.; Cooke, R.; Falk, R.T.; Gapstur, S.M.; Gaudet, M.M.; et al. Anthropometric and hormonal risk factors for male breast cancer: Male breast cancer pooling project results. *J. Natl. Cancer Inst.* **2014**, *106*, djt465. [[CrossRef](#)] [[PubMed](#)]
16. Brinton, L.A.; Key, T.J.; Kolonel, L.N.; Michels, K.B.; Sesso, H.D.; Ursin, G.; Eeden, S.K.V.D.; Wood, S.N.; Falk, R.T.; Parisi, D.; et al. Prediagnostic Sex Steroid Hormones in Relation to Male Breast Cancer Risk. *J. Clin. Oncol.* **2015**, *33*, 2041–2050. [[CrossRef](#)] [[PubMed](#)]
17. Arzanova, E.; Mayrovitz, H.N. Male Breast Cancer: Treatment Trends, Reported Outcomes, and Suggested Recommendations. *Cureus* **2021**, *13*, e18337. [[CrossRef](#)]
18. Hassett, M.J.; Somerfield, M.R.; Giordano, S.H. Management of Male Breast Cancer: ASCO Guideline Summary. *JCO Oncol. Pract.* **2020**, *16*, e839–e843. [[CrossRef](#)]
19. McClurg, D.P.; Urquhart, G.; McGoldrick, T.; Chatterji, S.; Miedzybrodzka, Z.; Speirs, V.; Elsberger, B. Analysis of the Clinical Advancements for. *Cancers* **2022**, *14*, 3175. [[CrossRef](#)]
20. Pemmaraju, N.; Munsell, M.F.; Hortobagyi, G.N.; Giordano, S.H. Retrospective review of male breast cancer patients: Analysis of tamoxifen-related side-effects. *Ann. Oncol.* **2012**, *23*, 1471–1474. [[CrossRef](#)]
21. Severson, T.M.; Kim, Y.; Joosten, S.E.P.; Schuurman, K.; van der Groep, P.; Moelans, C.B.; ter Hoeve, N.D.; Manson, Q.F.; Martens, J.W.; van Deurzen, C.H.M.; et al. Characterizing steroid hormone receptor chromatin binding landscapes in male and female breast cancer. *Nat. Commun.* **2018**, *9*, 482. [[CrossRef](#)]
22. Massarweh, S.A.; Sledge, G.W.; Miller, D.P.; McCullough, D.; Petkov, V.I.; Shak, S. Molecular Characterization and Mortality From Breast Cancer in Men. *J. Clin. Oncol.* **2018**, *36*, 1396–1404. [[CrossRef](#)] [[PubMed](#)]
23. Callari, M.; Cappelletti, V.; De Cecco, L.; Musella, V.; Miodini, P.; Veneroni, S.; Gariboldi, M.; Pierotti, M.A.; Daidone, M.G. Gene expression analysis reveals a different transcriptomic landscape in female and male breast cancer. *Breast Cancer Res. Treat.* **2011**, *127*, 601–610. [[CrossRef](#)] [[PubMed](#)]
24. Patel, A.P.; Tirosh, I.; Trombetta, J.J.; Shalek, A.K.; Gillespie, S.M.; Wakimoto, H.; Cahill, D.P.; Nahed, B.V.; Curry, W.T.; Martuza, R.L.; et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **2014**, *344*, 1396–1401. [[CrossRef](#)] [[PubMed](#)]
25. Lambrechts, D.; Wauters, E.; Boeckx, B.; Aibar, S.; Nittner, D.; Burton, O.; Bassez, A.; Decaluwé, H.; Pircher, A.; Van den Eynde, K.; et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.* **2018**, *24*, 1277–1289. [[CrossRef](#)]
26. Slyper, M.; Porter, C.B.M.; Ashenberg, O.; Waldman, J.; Drokhyansky, E.; Wakiro, I.; Smillie, C.; Smith-Rosario, G.; Wu, J.; Dionne, D.; et al. A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat. Med.* **2020**, *26*, 792–802. [[CrossRef](#)] [[PubMed](#)]
27. Davidson, S.; Efremova, M.; Riedel, A.; Mahata, B.; Pramanik, J.; Huuhtanen, J.; Kar, G.; Vento-Tormo, R.; Hagai, T.; Chen, X.; et al. Single-Cell RNA Sequencing Reveals a Dynamic Stromal Niche That Supports Tumor Growth. *Cell Rep.* **2020**, *31*, 107628. [[CrossRef](#)]
28. Kim, N.; Kim, H.K.; Lee, K.; Hong, Y.; Cho, J.H.; Choi, J.W.; Lee, J.-I.; Suh, Y.-L.; Ku, B.M.; Eum, H.H.; et al. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat. Commun.* **2020**, *11*, 2285. [[CrossRef](#)]
29. Cochran, D.R.; Campbell, K.R.; Greening, K.; Ho, G.C.; Hopkins, J.; Bui, M.; Douglas, J.M.; Sharlandjeva, V.; Munzur, A.D.; Lai, D.; et al. Single cell transcriptomes of normal endometrial derived organoids uncover novel cell type markers and cryptic differentiation of primary tumours. *J. Pathol.* **2020**, *252*, 201–214. [[CrossRef](#)]
30. Roadmap Epigenomics Consortium; Kundaje, A.; Meuleman, W.; Ernst, J.; Bilenky, M.; Yen, A.; Heravi-Moussavi, A.; Kheradpour, P.; Zhang, Z.; Wang, J.; et al. Integrative analysis of 111 reference human epigenomes. *Nature* **2015**, *518*, 317–330. [[CrossRef](#)]
31. Mansour, M.R.; Abraham, B.J.; Anders, L.; Berezovskaya, A.; Gutierrez, A.; Durbin, A.D.; Etchin, J.; Lawton, L.; Sallan, S.E.; Silverman, L.B.; et al. Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* **2014**, *346*, 1373–1377. [[CrossRef](#)]
32. Zhang, X.; Choi, P.S.; Francis, J.M.; Imielinski, M.; Watanabe, H.; Cherniack, A.D.; Meyerson, M. Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nat. Genet.* **2016**, *48*, 176–182. [[CrossRef](#)] [[PubMed](#)]
33. Roe, J.-S.; Hwang, C.-I.; Somerville, T.D.; Milazzo, J.P.; Lee, E.J.; Da Silva, B.; Maiorino, L.; Tiriack, H.; Young, C.M.; Miyabayashi, K.; et al. Enhancer Reprogramming Promotes Pancreatic Cancer Metastasis. *Cell* **2017**, *170*, 875–888.e20. [[CrossRef](#)] [[PubMed](#)]
34. Buenrostro, J.D.; Wu, B.; Litzenburger, U.M.; Ruff, D.; Gonzales, M.L.; Snyder, M.P.; Chang, H.Y.; Greenleaf, W.J. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **2015**, *523*, 486–490. [[CrossRef](#)] [[PubMed](#)]
35. Cusanovich, D.A.; Daza, R.; Adey, A.; Pliner, H.A.; Christiansen, L.; Gunderson, K.L.; Steemers, F.J.; Trapnell, C.; Shendure, J. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **2015**, *348*, 910–914. [[CrossRef](#)] [[PubMed](#)]

36. Granja, J.M.; Klemm, S.; McGinnis, L.M.; Kathiria, A.S.; Mezger, A.; Corces, M.R.; Parks, B.; Gars, E.; Liedtke, M.; Zheng, G.X.Y.; et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat. Biotechnol.* **2019**, *37*, 1458–1465. [CrossRef]
37. Regner, M.J.; Wisniewska, K.; Garcia-Recio, S.; Thennavan, A.; Mendez-Giraldez, R.; Malladi, V.S.; Hawkins, G.; Parker, J.S.; Perou, C.M.; Bae-Jump, V.L.; et al. A multi-omic single-cell landscape of human gynecologic malignancies. *Mol. Cell.* **2021**, *81*, 4924–4941.e10. [CrossRef]
38. Korsunsky, I.; Millard, N.; Fan, J.; Slowikowski, K.; Zhang, F.; Wei, K.; Baglaenko, Y.; Brenner, M.; Loh, P.-R.; Raychaudhuri, S. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **2019**, *16*, 1289–1296. [CrossRef]
39. Granja, J.M.; Corces, M.R.; Pierce, S.E.; Bagdatli, S.T.; Choudhry, H.; Chang, H.Y.; Greenleaf, W.J. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **2021**, *53*, 403–411. [CrossRef]
40. Hao, Y.; Hao, S.; Andersen-Nissen, E.; Mauck, W.M., 3rd; Zheng, S.; Butler, A.; Lee, M.J.; Wilk, A.J.; Darby, C.; Zager, M.; et al. Integrated analysis of multimodal single-cell data. *Cell* **2021**, *184*, 3573–3587.e29. [CrossRef]
41. Stuart, T.; Butler, A.; Hoffman, P.; Hafemeister, C.; Papalexi, E.; Mauck, W.M., III; Hao, Y.; Stoeckius, M.; Smibert, P.; Satija, R. Comprehensive Integration of Single-Cell Data. *Cell* **2019**, *177*, 1888–1902.e21. [CrossRef]
42. Tickle, T.I.; Georgescu, C.; Brown, M.; Haas, B. inferCNV of the Trinity CTAT Project. 2019. Available online: <https://github.com/broadinstitute/inferCNV> (accessed on 10 February 2022).
43. Pal, B.; Chen, Y.; Vaillant, F.; Capaldo, B.D.; Joyce, R.; Song, X.; Bryant, V.L.; Penington, J.S.; Di Stefano, L.; Ribera, N.T.; et al. A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. *EMBO J.* **2021**, *40*, e107333. [CrossRef]
44. Sahakyan, A.; Yang, Y.; Plath, K. The Role of Xist in X-Chromosome Dosage Compensation. *Trends Cell Biol.* **2018**, *28*, 999–1013. [CrossRef] [PubMed]
45. Liberzon, A.; Birger, C.; Thorvaldsdóttir, H.; Ghandi, M.; Mesirov, J.P.; Tamayo, P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **2015**, *1*, 417–425. [CrossRef]
46. Andrikopoulou, A.; Liontos, M.; Koutsoukos, K.; Dimopoulos, M.-A.; Zagouri, F. The emerging role of BET inhibitors in breast cancer. *Breast* **2020**, *53*, 152–163. [CrossRef] [PubMed]
47. Du, L.; Li, X.; Zhen, L.; Chen, W.; Mu, L.; Zhang, Y.; Song, A. Everolimus inhibits breast cancer cell growth through PI3K/AKT/mTOR signaling pathway. *Mol. Med. Rep.* **2018**, *17*, 7163–7169. [CrossRef]
48. Baselga, J.; Campone, M.; Piccart, M.; Burris, H.A., III; Rugo, H.S.; Sahmoud, T.; Noguchi, S.; Gnant, M.; Pritchard, K.I.; Lebrun, F.; et al. Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer. *N. Engl. J. Med.* **2012**, *366*, 520–529. [CrossRef] [PubMed]
49. The ENCODE Project Consortium; Moore, J.E.; Purcaro, M.J.; Pratt, H.E.; Epstein, C.B.; Shores, N.; Adrian, J.; Kawli, T.; Davis, C.A.; Dobin, A.; et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **2020**, *583*, 699–710. [CrossRef]
50. Abdelraouf, E.M.; Hussein, R.R.; Shaaban, A.H.; El-Sherief, H.A.; Embaby, A.S.; El-Aleem, S.A.A. Annexin A2 (AnxA2) association with the clinicopathological data in different breast cancer subtypes: A possible role for AnxA2 in tumor heterogeneity and cancer progression. *Life Sci.* **2022**, *308*, 120967. [CrossRef]
51. Gibbs, L.D.; Chaudhary, P.; Mansheim, K.; Hare, R.J.; Mantsch, R.A.; Vishwanatha, J.K. ANXA2 expression in African American triple-negative breast cancer patients. *Breast Cancer Res. Treat.* **2019**, *174*, 113–120. [CrossRef]
52. Weintraub, A.S.; Li, C.H.; Zamudio, A.V.; Sigova, A.A.; Hannett, N.M.; Day, D.S.; Abraham, B.J.; Cohen, M.A.; Nabet, B.; Buckley, D.L.; et al. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **2017**, *171*, 1573–1588.e28. [CrossRef]
53. Rafiei, S.; Tiedemann, K.; Tabariès, S.; Siegel, P.M.; Komarova, S.V. Peroxiredoxin 4: A novel secreted mediator of cancer induced osteoclastogenesis. *Cancer Lett.* **2015**, *361*, 262–270. [CrossRef]
54. Wang, G.; Zhong, W.-C.; Bi, Y.-H.; Tao, S.-Y.; Zhu, H.-X.; Xu, A.-M. The Prognosis Of Peroxiredoxin Family In Breast Cancer. *Cancer Manag. Res.* **2019**, *11*, 9685–9699. [CrossRef]
55. Hnisz, D.; Abraham, B.J.; Lee, T.I.; Lau, A.; Saint-André, V.; Sigova, A.A.; Hoke, H.A.; Young, R.A. Super-enhancers in the control of cell identity and disease. *Cell* **2013**, *155*, 934–947. [CrossRef]
56. Kelly, M.R.; Wisniewska, K.; Regner, M.J.; Lewis, M.W.; Perreault, A.A.; Davis, E.S.; Phanstiel, D.H.; Parker, J.S.; Franco, H.L. A multi-omic dissection of super-enhancer driven oncogenic gene expression programs in ovarian cancer. *Nat. Commun.* **2022**, *13*, 4247. [CrossRef]
57. Cassandri, M.; Butera, A.; Amelio, I.; Lena, A.M.; Montanaro, M.; Mauriello, A.; Anemona, L.; Candi, E.; Knight, R.A.; Agostini, M.; et al. ZNF750 represses breast cancer invasion via epigenetic control of prometastatic genes. *Oncogene* **2020**, *39*, 4331–4343. [CrossRef]
58. Islam, S.; Kitagawa, T.; Baron, B.; Abiko, Y.; Chiba, I.; Kuramitsu, Y. ITGA2, LAMB3, and LAMC2 may be the potential therapeutic targets in pancreatic ductal adenocarcinoma: An integrated bioinformatics analysis. *Sci. Rep.* **2021**, *11*, 10563. [CrossRef]
59. Betancur, P.A.; Abraham, B.J.; Yiu, Y.Y.; Willingham, S.B.; Khameneh, F.; Zarnegar, M.; Kuo, A.H.; McKenna, K.; Kojima, Y.; Leeper, N.J.; et al. A CD47-associated super-enhancer links pro-inflammatory signalling to CD47 upregulation in breast cancer. *Nat. Commun.* **2017**, *8*, 14802. [CrossRef] [PubMed]
60. Yuan, J.; Shi, X.; Chen, C.; He, H.; Liu, L.; Wu, J.; Yan, H. High expression of CD47 in triple negative breast cancer is associated with epithelial-mesenchymal transition and poor prognosis. *Oncol. Lett.* **2019**, *18*, 3249–3255. [CrossRef] [PubMed]

61. Wang, T.; Ma, S.; Qi, X.; Tang, X.; Cui, D.; Wang, Z.; Chi, J.; Li, P.; Zhai, B. Knockdown of the differentially expressed gene TNFRSF12A inhibits hepatocellular carcinoma cell proliferation and migration in vitro. *Mol. Med. Rep.* **2017**, *15*, 1172–1178. [[CrossRef](#)]
62. Xu, T.; Ruan, H.; Gao, S.; Liu, J.; Liu, Y.; Song, Z.; Cao, Q.; Wang, K.; Bao, L.; Liu, D.; et al. ISG20 serves as a potential biomarker and drives tumor progression in clear cell renal cell carcinoma. *Aging* **2020**, *12*, 1808–1827. [[CrossRef](#)]
63. Boutas, I.; Potiris, A.; Brenner, W.; Lebrecht, A.; Hasenburg, A.; Kalantaridou, S.; Schmidt, M. The expression of galectin-3 in breast cancer and its association with chemoresistance: A systematic review of the literature. *Arch. Gynecol. Obstet.* **2019**, *300*, 1113–1120. [[CrossRef](#)] [[PubMed](#)]
64. Zhang, F.; Zhang, L.; Zhang, B.; Wei, X.; Yang, Y.; Qi, R.Z.; Ying, G.; Zhang, N.; Niu, R. Anxa2 plays a critical role in enhanced invasiveness of the multidrug resistant human breast cancer cells. *J. Proteome Res.* **2009**, *8*, 5041–5047. [[CrossRef](#)] [[PubMed](#)]
65. Park, S.Y.; Lee, Y.; Park, J.; Kim, T.; Hong, S.; Jung, E.; Ju, Y.; Jeong, C.; Park, H.J.; Ko, G.H.; et al. PRDX4 overexpression is associated with poor prognosis in gastric cancer. *Oncol. Lett.* **2020**, *19*, 3522–3530. [[CrossRef](#)] [[PubMed](#)]
66. Zheng, J.; Guo, X.; Nakamura, Y.; Zhou, X.; Yamaguchi, R.; Zhang, J.; Ishigaki, Y.; Uramoto, H.; Yamada, S. Overexpression of PRDX4 Modulates Tumor Microenvironment and Promotes Urethane-Induced Lung Tumorigenesis. *Oxidative Med. Cell. Longev.* **2020**, *2020*, 8262730. [[CrossRef](#)]
67. Thandapani, P. Super-enhancers in cancer. *Pharmacol. Ther.* **2019**, *199*, 129–138. [[CrossRef](#)]
68. Noman, M.Z.; Van Moer, K.; Marani, V.; Gemmill, R.M.; Tranchevent, L.-C.; Azuaje, F.; Muller, A.; Chouaib, S.; Thiery, J.P.; Berchem, G.; et al. CD47 is a direct target of SNAI1 and ZEB1 and its blockade activates the phagocytosis of breast cancer cells undergoing EMT. *Oncoimmunology* **2018**, *7*, e1345415. [[CrossRef](#)]
69. Team, R.C. *R: A Language and Environment for Statistical Computing*; R Foundation: Vienna, Austria, 2020.
70. McGinnis, C.S.; Murrow, L.M.; Gartner, Z.J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* **2019**, *8*, 329–337.e4. [[CrossRef](#)]
71. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016.
72. Neftel, C.; Laffy, J.; Filbin, M.G.; Hara, T.; Shore, M.E.; Rahme, G.J.; Richman, A.R.; Silverbush, D.; Shaw, M.L.; Hebert, C.M.; et al. An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **2019**, *178*, 835–849.e21. [[CrossRef](#)]
73. Wu, S.Z.; Al-Eryani, G.; Roden, D.L.; Junankar, S.; Harvey, K.; Andersson, A.; Thennavan, A.; Wang, C.; Torpy, J.R.; Bartonicek, N.; et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat. Genet.* **2021**, *53*, 1334–1347. [[CrossRef](#)]
74. Aran, D.; Looney, A.P.; Liu, L.; Wu, E.; Fong, V.; Hsu, A.; Chak, S.; Naikawadi, R.P.; Wolters, P.J.; Abate, A.R.; et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **2019**, *20*, 163–172. [[CrossRef](#)]
75. Wu, T.; Hu, E.; Xu, S.; Chen, M.; Guo, P.; Dai, Z.; Feng, T.; Zhou, L.; Tang, W.; Zhan, L.; et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2021**, *2*, 100141. [[CrossRef](#)]
76. Scrucca, L.; Fop, M.; Murphy, T.B.; Raftery, A.E. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *R J.* **2016**, *8*, 289–317.
77. Satpathy, A.T.; Granja, J.M.; Yost, K.E.; Qi, Y.; Meschi, F.; McDermott, G.P.; Olsen, B.N.; Mumbach, M.R.; Pierce, S.E.; Corces, M.R.; et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **2019**, *37*, 925–936. [[CrossRef](#)]
78. Zhang, Y.; Liu, T.; Meyer, C.A.; Eeckhoutte, J.; Johnson, D.S.; Bernstein, B.E.; Nusbaum, C.; Myers, R.M.; Brown, M.; Li, W.; et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **2008**, *9*, R137. [[CrossRef](#)] [[PubMed](#)]
79. Liu, T. Use model-based Analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein-DNA interactions in embryonic stem cells. *Methods Mol. Biol.* **2014**, *1150*, 81–95. [[CrossRef](#)] [[PubMed](#)]
80. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **1995**, *57*, 289–300.
81. Zhu, L.J.; Gazin, C.; Lawson, N.D.; Pagès, H.; Lin, S.M.; Lapointe, D.S.; Green, M.R. ChIPpeakAnno: A Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinform.* **2010**, *11*, 237.
82. Lambert, S.A.; Yang, A.W.H.; Sasse, A.; Cowley, G.; Albu, M.; Caddick, M.X.; Morris, Q.D.; Weirauch, M.T.; Hughes, T.R. Similarity regression predicts evolution of transcription factor sequence specificity. *Nat. Genet.* **2019**, *51*, 981–989. [[CrossRef](#)]
83. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [[CrossRef](#)] [[PubMed](#)]
84. Danecek, P.; McCarthy, S.A. BCFTools/csqs: Haplotype-aware variant consequences. *Bioinformatics* **2017**, *33*, 2037–2039. [[CrossRef](#)]
85. Bailey, T.L.; Johnson, J.; Grant, C.E.; Noble, W.S. The MEME Suite. *Nucleic Acids Res.* **2015**, *43*, W39–W49. [[CrossRef](#)] [[PubMed](#)]
86. Fornes, O.; Castro-Mondragon, J.A.; Khan, A.; Van Der Lee, R.; Zhang, X.; Richmond, P.A.; Modi, B.P.; Correard, S.; Gheorghe, M.; Baranašić, D.; et al. JASPAR 2020: Update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **2020**, *48*, D87–D92. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.