

Article

# Unraveling the Genomic-Epigenomic Interaction Landscape in Triple Negative and Non-Triple Negative Breast Cancer

Jiande Wu <sup>1</sup>, Tarun Karthik Kumar Mamidi <sup>2</sup>, Lu Zhang <sup>3</sup> and Chindo Hicks <sup>1,\*</sup>

<sup>1</sup> Health Sciences Center, Department of Genetic, Louisiana State University School of Medicine, 533 Bolivar Street, New Orleans, LA 70112, USA; jwu2@lsuhsc.edu

<sup>2</sup> Center for Computational Genomics and Data Science, Departments of Pediatrics and Pathology, University of Alabama–Birmingham School of Medicine, Birmingham, AL 35233, USA; tmamidi@uab.edu

<sup>3</sup> Department of Public Health Sciences, Clemson University, 513 Edwards Hall, Clemson, SC 29634, USA; lz3@clemson.edu

\* Correspondence: chick3@lsuhsc.edu; Tel.: +1-504-568-2657

Received: 30 April 2020; Accepted: 5 June 2020; Published: 12 June 2020



**Abstract: Background:** The recent surge of next generation sequencing of breast cancer genomes has enabled development of comprehensive catalogues of somatic mutations and expanded the molecular classification of subtypes of breast cancer. However, somatic mutations and gene expression data have not been leveraged and integrated with epigenomic data to unravel the genomic-epigenomic interaction landscape of triple negative breast cancer (TNBC) and non-triple negative breast cancer (non-TNBC). **Methods:** We performed integrative data analysis combining somatic mutation, epigenomic and gene expression data from The Cancer Genome Atlas (TCGA) to unravel the possible oncogenic interactions between genomic and epigenomic variation in TNBC and non-TNBC. We hypothesized that within breast cancers, there are differences in somatic mutation, DNA methylation and gene expression signatures between TNBC and non-TNBC. We further hypothesized that genomic and epigenomic alterations affect gene regulatory networks and signaling pathways driving the two types of breast cancer. **Results:** The investigation revealed somatic mutated, epigenomic and gene expression signatures unique to TNBC and non-TNBC and signatures distinguishing the two types of breast cancer. In addition, the investigation revealed molecular networks and signaling pathways enriched for somatic mutations and epigenomic changes unique to each type of breast cancer. The most significant pathways for TNBC were: retinal biosynthesis, BAG2, LXR/RXR, EIF2 and P2Y purigenic receptor signaling pathways. The most significant pathways for non-TNBC were: UVB-induced MAPK, PCP, Apelin endothelial, Endoplasmatic reticulum stress and mechanisms of viral exit from host signaling Pathways. **Conclusion:** The investigation revealed integrated genomic, epigenomic and gene expression signatures and signing pathways unique to TNBC and non-TNBC, and a gene signature distinguishing the two types of breast cancer. The study demonstrates that integrative analysis of multi-omics data is a powerful approach for unravelling the genomic-epigenomic interaction landscape in TNBC and non-TNBC.

**Keywords:** gene expression; somatic; genomics; epigenomics; interactions; breast cancer

## 1. Introduction

Despite remarkable progress in screening and patient management, breast cancer remains the most commonly diagnosed non-skin cancer and the second leading cause of cancer related death among women in the US [1]. In 2019, 268,600 women were newly diagnosed with breast cancer, and 41,760 women died from the disease in the US [1]. Breast cancer is a heterogeneous disease

encompassing many types and subtypes. A majority of breast cancers are non-triple negative breast cancers (non-TNBC). These types of breast cancer are characterized by less aggressive clinical behavior, good prognosis, low recurrence and higher survival rates [2,3]. Importantly, these cancers respond to endocrine and targeted therapies [2,3].

However, a significant proportion of breast cancers ~15–20% are triple negative breast cancers (TNBC), the most aggressive and lethal form of breast cancer [2,3]. TNBC are defined as tumors that lack expression of the estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor (HER-2) amplification [2,3]. TNBC is characterized by poorer prognosis, higher recurrence rates and worse overall survival rates than non-TNBC [2,3]. Most notably, unlike non-TNBC, currently there are no effective targeted therapies for TNBC, cytotoxic chemotherapy remains the only effective therapeutic modality [2,3]. Accumulating evidence from both epidemiologic and clinical studies has revealed that TNBC has worse outcomes and survival rates than non-TNBC [4–11]. However, the molecular mechanisms underlying the etiological and biological differences between the two types of breast cancer have not been well characterized. Thus, a key knowledge gap and critical unmet medical need pivot around: (i) understanding the molecular mechanisms underlying the biological differences between TNBC and non-TNBC; (ii) the discovery of clinically actionable molecular markers and potential targets for the development of novel therapeutics unique to each type of breast; and (iii) the identification of biomarkers distinguishing the two types of breast cancer that could be used to identify and stratify patients to guide treatment decisions.

Our understanding of the molecular mechanisms underlying the biological differences between TNBC and non-TNBC and how best to treat them is hampered by the biological and etiological complexity of the two diseases. Both types of breast cancer are heterogeneous diseases, each involving many subtypes, and their development and progression involves a complex interplay between genomic and epigenomic alterations and many environmental perturbations. These complex arrays of interacting factors likely affect entire molecular networks and signaling pathways, which in turn drive the two types of breast cancer. Integrative data analysis combining genomic, epigenomic and gene expression data has the promise to increase our understanding of the biological mechanisms driving the two types of breast cancer, and to identify clinically actionable molecular markers and targets to guide therapeutic decisions.

Advances in microarray technology have enabled the molecular classification of types and subtypes of breast cancer using transcription profiling [12–14]. Evidence from these primary studies have revealed differences in the molecular portraits of TNBC and non-TNBC [12–14]. Transcription profiling has enabled development of two prognostic signatures, the PAM50 [15,16] and the MammaPrint [17,18], which have proven useful for prognostic purposes in the clinic [15–18]. However, although these primary analyses have made great strides in deciphering the molecular basis of breast cancer and enabled development of prognostic signatures, they have been unsuccessful in determining how genomic and epigenomic factors interact and cooperate to drive each type of breast cancer.

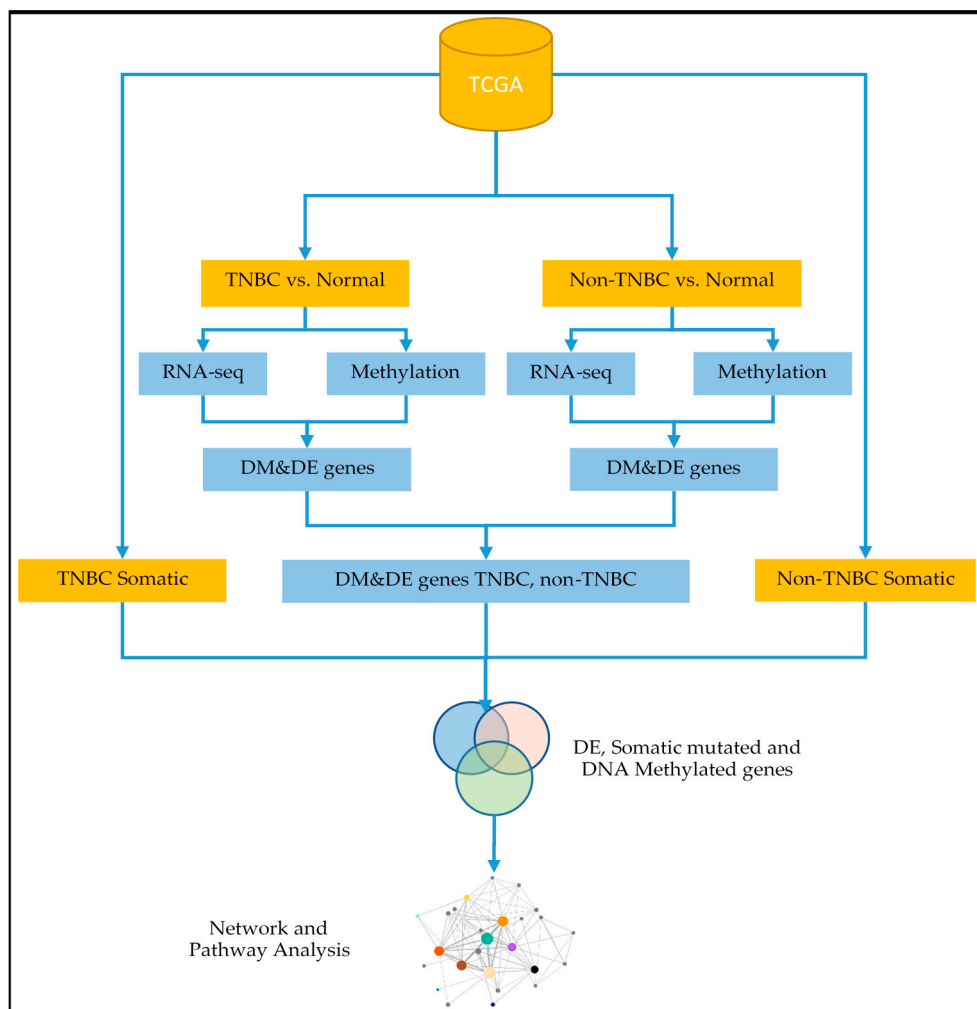
Over the last decade, rapid development in next generation sequencing technologies has brought considerable advances in large-scale sequencing of cancer genomes, including breast cancer [19–21]. These advances have enabled development of comprehensive catalogues of somatic mutations and epigenomic signatures of breast cancer. In addition, these advances have led to an expanded molecular classification and increased our understanding of the molecular taxonomy of breast cancer. Traditionally, the analyses of somatic, epigenomic and gene expression data in different types of breast cancer have been conducted as separate research endeavors. Recently, we reported a novel approach integrating somatic mutation with DNA methylation information using gene expression data in TNBC [22]. This investigation showed that somatic mutated genes harbor epigenomic alterations, and that epigenetic changes affect gene expression [22]. However, to date, information on somatic mutations has not been leveraged and integrated with DNA methylation data to unravel the possible oncogenic interactions between genomic and epigenomic alterations in TNBC and non-TNBC. The objectives of this study were to: (i) define integrated genomic, epigenomic and gene expression profiling signatures

unique to TNBC from non-TNBC; (ii) discover a signature distinguishing the two types of breast cancer; and (iii) map the possible oncogenic interactions between genomic and epigenomic alterations in TNBC and non-TNBC, using network and pathways analysis. Our working hypothesis was that within breast cancers, there are differences in somatic mutation, DNA methylation and gene expression signatures between TNBC and non-TNBC. We further hypothesized that genomic and epigenomic alterations affect gene regulatory networks and signaling pathways driving the two types of breast cancer. We addressed these hypotheses using an integrative genomics data analysis approach, combining information on somatic, epigenomic and gene expression data on TNBC and non-TNBC from The Cancer Genome Atlas (TCGA).

## 2. Materials and Methods

### 2.1. Overall Project Design and Execution Strategy

The overall project design and workflow for integrative analyses strategies, integrating somatic, DNA methylation and gene expression data, is presented in Figure 1.



**Figure 1.** Project design and integrated analysis workflow integrating somatic mutation, DNA methylation and gene expression data leveraged with network, pathway and functional analysis to unravel the genomic-epigenomic interaction landscape in triple negative breast cancer (TNBC) and non-triple negative breast cancer (non-TNBC). DM indicates differentially methylated, DE indicates differentially expressed.

## 2.2. Sources of Genomic and Epigenomic Data

The TCGA data management team has assembled genomic, epigenomic and mutation data, along with clinical information, and has made it available to the cancer research community [19,20] via the Genomics Data Commons [23]. For this study, we used gene expression, DNA methylation (array-based) and somatic mutation data from the TCGA [19]. The data sets were downloaded from the Genomics Data Commons data portal using the data transfer tool [23]. The original gene expression data set was generated using RNA-Seq and consisted of 828 samples, distributed as follows: patients diagnosed with TNBC tumors  $n = 110$  samples, patients diagnosed with non-TNBC tumors  $n = 605$  samples and controls  $n = 113$  samples (Table 1). The original gene expression data matrix included 60,484 probes (Table 1).

**Table 1.** Characteristics and distribution of samples, probes and gene symbols for the original data sets on gene expression, somatic mutation and DNA methylation for triple negative breast cancer (TNBC) and non-triple negative breast cancer (non-TNBC) used in this study.

Data Type	TNBC		Non-TNBC		Control
	Genes or Probes	Tumor Samples	Genes or Probes	Tumor Samples	Samples
Gene Expression	60,484 probes	110	60,484 probes	605	113
Methylation	485,578 probes	83	485,578 probes	597	83
Somatic Mutated	7659 genes	110	16,752 genes	605	-

The original DNA methylation (array-based) data set consisted of 763 samples, distributed as follows: patients diagnosed with TNBC tumors  $n = 83$  samples, patients diagnosed with non-TNBC tumors  $n = 597$  samples and controls  $n = 83$  samples (Table 1). DNA methylation data was generated using Illumina HumanMethylation450 BeadChip (Illumina Inc. San Diego, CA, USA). This chip has been widely used for quantifying DNA methylation and has been validated [24]. The DNA methylation data matrix included 485,578 probes (Table 1). We processed and performed data quality control (QC) using the HumanMethylation450 evaluation and processing methods [25,26], implemented in the DNA methylation data analysis pipeline developed in our lab [22]. The data was further processed to remove batch effects prior to analysis [22,26]. Considering the convolution of biological and technological variability and the signal bias between the Infinium I and II probe design types, the data was further quality controlled to correct for type of probe design [27], and to eliminate bias attributable to systematic errors consistent with the Illumina data analysis protocol [25,26]. We used the quantile normalization implemented in R software package [22,24–27] to normalize the data and calibrated for batch effects consistent with the Illumina data analysis protocol [25–27].

The DNA methylation data was derived from the same patient population as gene expression and somatic mutation data. Therefore, following processing and quality control of disparate data sets, we optimized the data matrix for integrated analysis by identifying samples with all three pieces of information. To address this issue, we used clinical information to link the methylation samples with the gene expression samples. After this processing step, we used 83 patients diagnosed with TNBC tumors, 597 patients diagnosed with non-TNBC tumors and 83 controls (Table 1) in the analysis. We then used a barcode structure provided by TCGA to integrate patient-based clinical data with sample-based somatic mutation data, and processed somatic mutation data to identify the number of genes that contain somatic mutations and the number of somatic mutation events for each gene across samples within the tumor and across tumors. In this processing step, we created a catalog of 7659 somatic mutant genes in TNBC and 16,752 somatic mutant genes in non-TNBC for analysis (Table 1). Supplementary Table S1 presents a comprehensive list of TNBC and non-TNBC somatic mutation genes and the number of somatic mutation events for each gene.

### 2.3. Bioinformatics Analysis of Disparate Omics Data

Figure 1 shows the data processing and analysis workflow. In this analysis, we focused on gene expression and DNA methylation data, analyzing them separately for each type of breast cancer. Note that only samples with both gene expression and DNA methylation were used in these analyses. We performed QC and noise reduction on the original gene expression data matrix to remove rows with insufficient information or missing data. The filtering was imposed such that each row had at least  $\geq 30\%$  data. After this QC step, the data matrix for TNBC reduced to 166 samples (83 TNBC tumors samples and 83 controls) with 28,084 probes used in the analysis. Likewise, using the same data processing and QC for non-TNBC reduced the data matrix to 680 samples distributed as (597 non-TNBCs tumor samples and 83 controls) with 27,218 probes used in the analysis. The quality-controlled data sets were normalized using the LIMMA Bioconductor package implemented in R [28]. The probe IDs were matched with gene symbols using the Ensemble database [29]. Gene expression data from RNA-Seq was quantified in Transcripts Per Kilobase Million (TPM), and was first  $\log_2(\text{TPM} + 1)$  transformed using the standard protocol for RNA sequence data normalization methods [30].

Using quality controlled normalized data sets, we compared gene expression levels between TNBC tumors and control samples, to discover a signature of significantly differentially expressed genes in TNBC. Likewise, we compared gene expression levels between non-TNBC tumors and control samples, to discover a signature of significantly differentially expressed genes in non-TNBC. Additionally, we performed supervised analysis comparing gene expression levels between TNBC and non-TNBC, to identify a signature of significantly differentially expressed genes distinguishing the two types of breast cancer. These robust and unbiased analyses enabled the discovery of all differentially expressed genes, including somatic mutated ones. Gene expression analysis was performed using the Bioconductor R-package LIMMA [28]. Throughout these analyses, we used the false discovery rate (FDR) method [31] to adjust the  $p$ -value for multiple hypothesis testing. In addition, for each analysis, we calculated the  $\log_2$  Fold Change (Log<sub>2</sub> FC), defined as the median of gene expressed minus the gene expression value for each gene.

For each type of breast cancer, genes were ranked on adjusted  $p$ -value, FDR and LogFC, and significantly ( $p < 0.05$ ) differentially expressed genes were identified and selected. Note that significant differential expression provides proof of association of these genes with the types of breast cancer. The discovered significantly differentially expressed genes were evaluated for the presence of somatic mutations to discover signatures of somatic mutated genes, which are significantly differentially expressed in each type of breast cancer and a signature of somatic mutated genes distinguishing the two types of breast cancer. Somatic mutated genes were further quantitatively assessed for the number of mutation events. A gene was considered highly mutated if it contained at least  $\geq 3$  mutation events. A gene was considered differentially mutated if it was mutated only in one type of breast cancer, or if the mutations found in one type of breast cancer were not found in the other type.

Likewise for DNA methylation data, we performed QC to remove batch effects [22] and noise reduction on the original data set to remove rows with insufficient information or missing data, using the same filtering threshold of each row having at least  $\geq 30\%$  data. After this QC step, the data matrix for TNBC reduced to 166 samples (83 TNBC tumors samples and 83 controls), with 383,119 probes used in the analysis. Using the same data processing and QC for non-TNBC reduced the data matrix to 680 samples distributed as (597 non-TNBCs tumor samples and 83 controls), with 365,947 probes used in the analysis. The quality-controlled data sets were normalized using quantile normalization applied to the LIMMA Bioconductor package implemented in R [22,27,28].

Using the quality controlled normalized DNA methylation data set, we compared DNA methylation profiles using beta-values between TNBC tumor and control samples and between non-TNBC tumor and control samples, to discover signatures of significantly ( $p < 0.05$ ) differentially methylated genes associated with each type of breast cancer [22]. These analyses were accomplished using the methods for analysis of DNA methylation data in the Illumina protocol [24–26], implemented in our pipeline [22]. In addition, we performed supervised analysis comparing DNA methylation

profiles between TNBC and non-TNBC, to identify a signature of significantly differentially methylated genes between the two types of breast cancer. We used the LIMMA package to identify differential methylated CpG sites [22,28]. For all the analyses we used the FDR [31] to control for multiple hypothesis testing. Genes were ranked on adjusted  $p$ -values, FDR and LogFC. We used the Ensemble Biomart database to link CpG sites with gene symbols [22,32]. Significantly differentially methylated genes were further quantitatively evaluated for the number of differentially CpG sites. A gene was considered highly differentially methylated if it contained  $\geq 3$  methylation events. A gene was considered differentially methylated if it was aberrantly methylated only in one type of breast cancer, or if the CpG sites found in one type of breast cancer were not found in the other type.

For each type of breast cancer, genes were ranked on adjusted differential methylation  $p$ -values, number of CpG sites per gene, FDR and LogFC. The discovered significantly differentially methylated genes were evaluated for the presence of somatic mutations, to discover the signatures of somatic mutated genes that are significantly differentially methylated in each type of breast cancer, and a signature of differentially methylated somatic mutated genes, distinguishing the two types of breast cancer. Somatic mutated genes were further quantitatively assessed to determine the number of mutation events. The same criteria as in gene expression was used for identifying highly mutated and differentially mutated genes.

#### 2.4. Integrative Analysis of Diverse Omics Data

We used a systems level integration to investigate potential oncogenic interactions between genomic and epigenomic alterations in TNBC and non-TNBC. The interplay between genomic and epigenomic alterations may happen both in *cis* and in *trans*, so we performed multiple levels of integration. At level 1 integration, we integrated information on differentially expressed genes with information on differentially methylated genes in TNBC and non-TNBC. Here, we sought to: (i) discover the signatures of differentially expressed genes, which are also differentially methylated (integrated signature) and uniquely associated with each type of breast cancer; and (ii) to discover a signature of significantly differentially expressed genes which were also significantly differentially methylated distinguishing the two types of breast cancer. At level 2 integration, we combined information on differentially methylated, expressed and somatic mutated in TNBC and non-TNBC. Here; we sought to: (i) discover the integrated signatures of differentially expressed genes, which are also differentially methylated and differentially mutated, uniquely associated with each type of breast cancer; and (ii) to discover an integrated signature of significantly differentially expressed genes, which are also differentially methylated and differentially mutated, distinguishing the two types of breast cancer. For each integrated signature, we conducted a quantitative assessment of somatic mutations per gene, within and between the two types of breast cancer, to assess the uniqueness and the differences in mutation burden between TNBC and non-TNBC.

The final levels of integrative analysis involved network (level 3), pathway (level 4) and functional (level 5) analysis of genes containing somatic and epigenomic alterations. Here, we sought to unravel the possible oncogenic interactions between genomic and epigenomic in TNBC and non-TNBC. We performed network and pathway prediction using the Ingenuity Pathway Analysis (IPA) Software (Qiagen Inc., Hilden, German) [33] to identify the molecular networks and signaling pathways enriched for somatic and epigenomic alterations in each type of breast cancer. Somatic mutated genes, which were differentially methylated and transcriptionally associated with each type of breast cancer were mapped onto networks and canonical pathways using IPA. For network analysis, a Z-score was computed, whereas for pathway prediction, the  $-\log(p\text{-values})$  were computed and used as the metrics for correctly assigning the genes to the correct networks and pathways, respectively. The FDR was used to correct for type I and type II errors, to ensure the reliability of correctly predicting the pathways. We performed GO [34] analysis using the IPA database to functionally characterize the discovered genes into the molecular functions, biological process and cellular components in which they are involved.

### 3. Results

#### 3.1. Gene Expression and DNA Methylation Signatures Associated with TNBC and Non-TNBC

We performed whole transcriptome and whole methylome analysis to identify gene expression and DNA methylation profiling signatures uniquely associated with TNBC and non-TNBC and signatures distinguishing the two types of breast cancer. We hypothesized that the molecular perturbation and aberrant DNA methylation in tumors from patients diagnosed with TNBC or non-TNBC and control samples could lead to measurable changes distinguishing patients diagnosed with TNBC tumors from control samples, and patients diagnosed with non-TNBC tumors from control samples. To address these hypotheses, we performed whole transcriptome and whole methylome analysis comparing gene expression levels and DNA methylation profiles between tumor and control samples for TNBC and non-TNBC separately.

Whole transcriptome analysis comparing gene expression levels between tumor and control samples in TNBC produced a signature of 15,404 significantly ( $p < 0.05$ ) differentially expressed genes. Comparing gene expression levels between tumor samples and controls in non-TNBC tumors produced a signature of 16,124 significantly ( $p < 0.05$ ) differentially expressed genes. Complete lists of the 15,404 genes associated with TNBC and the 16,124 genes associated with non-TNBC along with estimates of  $p$ -values, FC indicative of the level of association, as well as their direction of change (up or down regulated) are presented in supplementary Table SE.

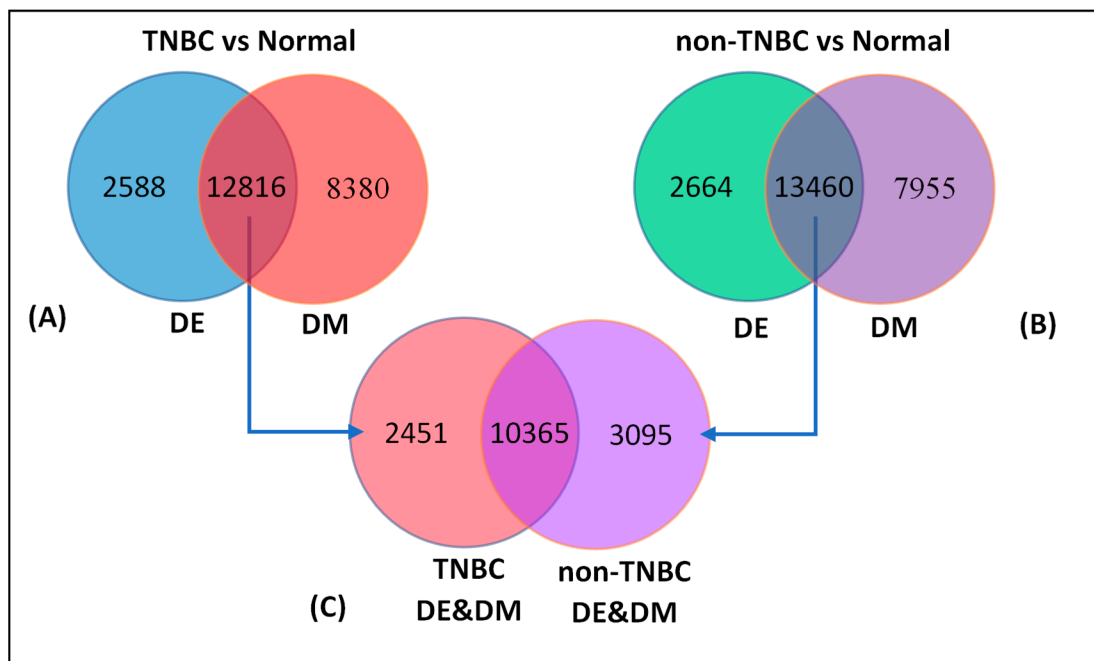
Whole methylome analysis comparing DNA methylation profiles between tumor and control samples in TNBC produced a signature of 21,196 significantly ( $p < 0.05$ ) differentially methylated genes. Comparing DNA methylation profiles between patients diagnosed with non-TNBC and controls revealed a signature of 21,415 significantly ( $p < 0.05$ ) differentially methylated genes. Complete lists of the 21,196 genes significantly differentially methylated in TNBC and the 21,415 genes significantly differentially methylated in non-TNBC, along with estimates of  $p$ -values, FC indicative of the level of association, as well as their direction of change (hyper or hypo methylated), are presented in supplementary Table SM.

#### 3.2. Discovery of Integrated DNA Methylation and Gene Expression Signatures

To discover the signatures of significantly ( $p < 0.05$ ) differentially expressed genes, which were also significantly ( $p < 0.05$ ) differentially methylated (herein called integrated signatures) unique to TNBC and non-TNBC, or genomically and epigenomically altered in both types of breast cancer, we performed analysis combining information on significantly ( $p < 0.05$ ) differentially expressed genes and significantly ( $p < 0.05$ ) differentially methylated genes in each type of breast cancer, compared to controls. For TNBC, we combined the 15,404 significantly ( $p < 0.05$ ) differentially expressed genes with the 21,196 significantly ( $p < 0.05$ ) differentially methylated genes. Similarly, for non-TNBC, we combined the 16,124 significantly ( $p < 0.05$ ) differentially expressed genes with the 21,415 significantly ( $p < 0.05$ ) differentially methylated genes. Genes were then sorted by gene symbols to identify the integrated DNA methylation and gene expression signatures unique to each type of breast cancer, and signatures that overlapped the two types of breast cancer.

The results of these analyses are summarized in Figure 2. For TNBC, we discovered an integrated DNA methylation and gene expression signature of 12,816 genes. In addition, 2588 genes were significantly ( $p < 0.05$ ) differentially expressed, but not differentially methylated, whereas 8380 genes were significantly ( $p < 0.05$ ) differentially methylated, but not differentially expressed (Figure 2A). For non-TNBC, we discovered an integrated DNA methylation and gene expression signature of 13,460 genes. In addition, we discovered a signature of 2664 significantly ( $p < 0.05$ ) differentially expressed genes, which were not aberrantly methylated and a signature of 7955 significantly ( $p < 0.05$ ) differentially methylated genes, which were not differentially expressed (Figure 2B). These findings confirmed our hypothesis that, within each type of breast cancer, there were genes that were significantly altered in both the transcriptome and the methylome. Supplementary Table S2A,B shows complete lists

of all the 12,816 genes in TNBC and 13,460 genes in non-TNBC, respectively, which were significantly ( $p < 0.05$ ) altered in both the transcriptome and the methylome.



**Figure 2.** Number of genes significantly ( $p < 0.05$ ) differentially expressed or significantly ( $p < 0.05$ ) differentially methylated in each type of breast cancer (outside intersections), and genes significantly ( $p < 0.05$ ) differentially expressed and significantly ( $p < 0.05$ ) differentially methylated in each type of breast cancer (in the intersection) are shown in (A) for TNBC and (B) for non-TNBC, respectively. Figure (C) shows numbers of significantly ( $p < 0.05$ ) differentially expressed genes that are significantly ( $p < 0.05$ ) differentially methylated unique to each type of breast cancer (outside the interaction), and genes overlapping the two types of breast cancer (see intersections). Please note, as indicated by the arrows, that the numbers in (C) are derived from the number intersections of (A) and (B).

Having discovered the signatures of genes altered in both the transcriptome and the methylome in each type of breast cancer (Figure 2A,B intersections), we performed an additional investigation to discover the signatures of genes unique to each type of breast cancer, and a signature of genes genomically and epigenomically altered in both types of breast cancer. We addressed this question by combining the 12,816 genes from TNBC (Figure 2A) with the 13,460 genes from non-TNBC (Figure 2B) with both alterations, and sorted them by gene symbol, expression and methylation  $p$ -values. The results of this investigation are presented in Figure 2C. The analysis revealed a signature of 10,365 genes genomically and epigenomically altered in TNBC and non-TNBC. In addition, the analysis revealed a signature of 2451 genes genomically and epigenomically altered in TNBC only, and a signature of 3095 genes altered in non-TNBC only (Figure 2C). A complete list of all the 10,365 significantly ( $p < 0.05$ ) differentially expressed genes, which were also significantly ( $p < 0.05$ ) differentially methylated in TNBC and non-TNBC, is presented in supplementary Table S2C. These analyses suggest that for a subset of genes, molecular perturbation and epigenomic changes occur in both types of breast cancer, but at different rates.

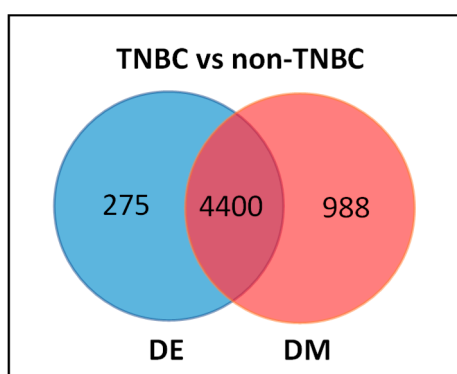
### 3.3. Discovery of Epigenomic and Gene Expression Signatures Distinguishing TNBC from Non-TNBC

A critical need in understanding the biology of breast cancer is characterizing the molecular differences between TNBC and non-TNBC tumors. We addressed this issue by testing the hypothesis that genomic and epigenomic (or a combination thereof) alterations differ between TNBC and non-TNBC. Here, we reasoned that, because of the etiological and biological differences between TNBC



and non-TNBC, the genomic and epigenomic mechanisms driving the two types of breast cancer may be different. To test this hypothesis, we created and used a new data set containing the 10,365 genes genomically and epigenomically altered in both TNBC and non-TNBC derived from (Figure 2C). We compared their expression levels and DNA methylation profiles between TNBC and non-TNBC.

The results of this investigation are shown in a Venn diagrams in Figure 3. We discovered a signature of 4675 significantly ( $p < 0.05$ ) differentially altered genes, and a signature of 5388 significantly ( $p < 0.05$ ) differentially methylated genes distinguishing the two types of breast cancer. Interestingly, we also discovered a signature of 4400 genes significantly ( $p < 0.05$ ) differentially in the transcriptome, which were also significantly ( $p < 0.05$ ) differentially methylated, distinguishing TNBC from non-TNBC (Figure 3, intersection). In addition, the analysis revealed a signature of 275 significantly ( $p < 0.05$ ) differentially expressed genes without DNA methylation, and a signature of 988 significantly ( $p < 0.05$ ) differentially methylated genes not differentially expressed (Figure 3) between the two types of breast cancer. This confirmed our hypothesis that an integrative analysis of genomic and epigenomic data could lead to the discovery of a signature of genes altered both in the transcriptome and the methylome, and distinguishing the two types of breast cancer.



**Figure 3.** Venn diagram showing the distribution of genes containing both differentially expressed (DE) and differentially methylated (DM), showing the 4400 genes distinguishing TNBC from non-TNBC (in the intersection), 275 differentially expressed (DE) only, and the 988 differentially methylated (DM) only between TNBC and non-TNBC.

To further explore the patterns of gene expression and DNA methylation, we evaluated the 4400 genes (Figure 3). The results showing the distribution of DNA methylation for the top list of genes that were highly significantly differentially altered in the transcriptome and the methylome between TNBC and non-TNBC are presented in Table 2. Only the genes with a high frequency of DNA methylation (DM) sites per gene are presented in the Table. A full list of all the 4400 genes significantly differentially altered in both the transcriptome and the methylome distinguishing the two types of breast cancer is presented in supplementary Table S3A. A full list of all the 5663 genes with a binary pattern of presence and absence is presented in supplementary Table S3B.

Overall, there was significant variation in patterns of DNA methylation, as measured by the distribution of significantly differentially methylated CpG sites and gene expression levels within and between the two subtypes of breast cancer. The observed variation in patterns of gene expression and DNA methylation profiles can be explained partially by the heterogeneity of these diseases. Both TNBC and non-TNBC are inherently heterogeneous types of breast cancer, each comprised of many subtypes of breast cancer. Under such conditions, the observed outcome should be expected. Due to the lack of information on the subtypes of breast cancer in the data sets used in this investigation, we did not address the issue of subtypes, as it was beyond the scope of this study.

**Table 2.** List of top 30 most highly significantly differentially expressed genes, which were also significantly differentially methylated distinguishing TNBC from non-TNBC, along with number of DNA methylation (DM) sites per gene and estimates of differential expression and differential methylation *p*-values. All *p*-values were adjusted for the false discovery rates.

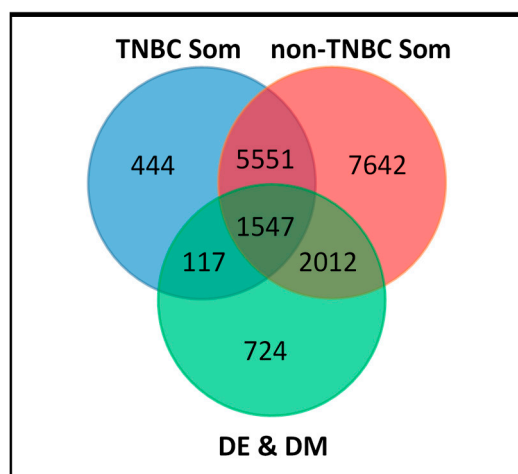
Gene_Symbol	Chromosome Position	EXP-Value	DM_Sites	DM <i>p</i> -Value
<i>PTPRN2</i>	7q36.3	$1.08 \times 10^{-15}$	487	$1.69 \times 10^{-19}$
<i>PCDHGA1</i>	5q31	$4.49 \times 10^{-2}$	242	$1.36 \times 10^{-22}$
<i>MSLN</i>	16p13.3	$2.86 \times 10^{-8}$	165	$1.50 \times 10^{-24}$
<i>CBFA2T3</i>	16q24.3	$8.05 \times 10^{-12}$	146	$4.07 \times 10^{-25}$
<i>CUX1</i>	7q22.1	$1.01 \times 10^{-3}$	120	$8.13 \times 10^{-28}$
<i>ANKRD11</i>	16q24.3	$2.92 \times 10^{-5}$	114	$3.40 \times 10^{-21}$
<i>NCOR2</i>	12q24.31	$1.97 \times 10^{-2}$	112	$1.07 \times 10^{-27}$
<i>LMF1</i>	16p13.3	$1.62 \times 10^{-14}$	111	$9.23 \times 10^{-27}$
<i>ABR</i>	17p13.3	$1.32 \times 10^{-3}$	103	$1.39 \times 10^{-24}$
<i>SLC45A4</i>	8q24.3	$1.12 \times 10^{-4}$	103	$3.29 \times 10^{-31}$
<i>SYNGAP1</i>	6p21.32	$5.78 \times 10^{-6}$	99	$2.34 \times 10^{-20}$
<i>TP73</i>	1p36.32	$2.26 \times 10^{-6}$	96	$2.27 \times 10^{-25}$
<i>PRKAR1B</i>	7p22.3	$9.45 \times 10^{-3}$	95	$3.32 \times 10^{-27}$
<i>ADAMTS2</i>	5q35.3	$1.47 \times 10^{-2}$	86	$1.07 \times 10^{-28}$
<i>PRRT1</i>	6p21.32	$3.99 \times 10^{-7}$	86	$1.63 \times 10^{-26}$
<i>PCDHA2</i>	5q31.3	$1.09 \times 10^{-3}$	83	$2.04 \times 10^{-10}$
<i>ANK1</i>	8p11.21	$5.84 \times 10^{-3}$	82	$1.75 \times 10^{-25}$
<i>TBC1D16</i>	17q25.3	$5.65 \times 10^{-3}$	81	$3.31 \times 10^{-31}$
<i>HOOK2</i>	19p13.13	$1.31 \times 10^{-4}$	79	$3.86 \times 10^{-23}$
<i>COL23A1</i>	5q35.3	$2.15 \times 10^{-5}$	75	$1.10 \times 10^{-34}$
<i>NEFATC1</i>	18q23	$1.38 \times 10^{-2}$	75	$9.25 \times 10^{-33}$
<i>FO XK1</i>	7p22.1	$2.74 \times 10^{-5}$	73	$4.24 \times 10^{-33}$
<i>BCOR</i>	Xp11.4	$1.09 \times 10^{-7}$	72	$2.95 \times 10^{-21}$
<i>COL4A2</i>	13q34	$3.03 \times 10^{-2}$	72	$5.02 \times 10^{-25}$
<i>PCDHA4</i>	5q31.3	$2.16 \times 10^{-5}$	72	$2.04 \times 10^{-10}$
<i>RGS12</i>	4p16.3	$3.48 \times 10^{-8}$	71	$1.21 \times 10^{-31}$
<i>RGL2</i>	6p21.32	$1.31 \times 10^{-14}$	69	$4.34 \times 10^{-29}$
<i>DNAH17</i>	17q25.3	$1.22 \times 10^{-4}$	68	$1.74 \times 10^{-26}$
<i>CTBP2</i>	10q26.13	$1.51 \times 10^{-4}$	67	$6.16 \times 10^{-26}$
<i>HOXC4</i>	12q13.13	$1.60 \times 10^{-9}$	67	$3.63 \times 10^{-11}$

### 3.4. Integration of Somatic and Epigenomic Variation

Having discovered an integrated signature of genes altered both in the transcriptome and the methylome distinguishing TNBC from non-TNBC, we conducted further investigations to integrate somatic mutation information. We hypothesized that genes differentially altered in the transcriptome and the methylome harbor somatic mutations. We further hypothesized that there are differences in somatic mutation burden and epigenetic alterations between the two types of breast cancer. Here, we sought to discover the somatic, epigenomic and gene expression signatures unique to TNBC and non-TNBC, and a signature distinguishing the two types of breast cancer. We addressed this issue as follows: first, we evaluated differentially expressed and differentially methylated genes for the presence of somatic mutations, using the 7659 somatic mutated genes in TNBC and the 16,752 genes somatic mutated genes in non-TNBC. We also evaluated the 4400 genes altered in the transcriptome and the methylome distinguishing TNBC from non-TNBC for the presence of mutations, and for differences in mutation burden between the two types of breast cancer. Finally, we performed a quantitative assessment of the differences in somatic mutations and epigenomic alterations between TNBC and non-TNBC.

The results of these analyses are summarized in a Venn diagram in Figure 4. Out of the 7659 genes somatic mutated in TNBC, 444 genes contained somatic mutations in TNBC only, and were neither differentially methylated nor differentially expressed between the two types of breast cancer

(Figure 4). Out of the 16,752 genes somatic mutated in non-TNBC, 7642 contained somatic mutations in non-TNBC only, and were neither differentially methylated nor differentially expressed between the two types of breast cancer (Figure 4). However, a complete evaluation of all somatic mutated genes for genomic and epigenomic alterations revealed a total of 5551 genes containing somatic mutations in both types of breast cancer, and were neither differentially methylated nor differentially expressed between the two types of breast cancer (Figure 4).



**Figure 4.** Somatic and epigenomic altered gene signatures unique to TNBC and non-TNBC, signatures common to both types of breast cancer and signatures distinguishing them.

Quantitative evaluation of the somatic mutated genes against the 4400 genes altered in both the transcriptome and the methylome distinguishing TNBC from non-TNBC revealed three gene signatures. Among them were the 117 gene signature containing both somatic and epigenomic alterations unique to TNBC, the 2012 gene signature containing both somatic and epigenomic alterations unique to non-TNBC and the 1547 gene signature containing both somatic and epigenomic alterations, which overlapped the two types of breast cancer (Figure 4). In addition, we discovered a 724 gene signature containing epigenomic alterations without somatic mutations (Figure 4). There was significant variation in both somatic mutation and DNA methylation profiles within and between the two types of breast cancer. As noted earlier in this report, this was expected, because breast cancer is inherently a heterogeneous disease, thus, the observed outcome was expected.

To further assess the patterns of gene expression, DNA methylation and somatic mutation profiles, we evaluated the genes containing all the three variations. Table 3 shows subsets of the top somatic mutated and epigenomically altered genes from the 117 genes unique to TNBC and the 2012 genes unique to non-TNBC, which distinguished the two types of breast cancer. The distributions of methylation sites, somatic mutation events and adjusted estimates of differential expression and differential methylation *p*-values are also presented in Table 3. Among the 117 genes unique to TNBC, the number of somatic mutation events per gene ranged from one to three, whereas the number of DNA methylation sites or CpG sites ranged from one to 63. Among the 2012 genes unique to non-TNBC, the number of somatic mutation events per gene ranged from one to 33, whereas the number of DNA methylation sites or CpG sites ranged from one to 165. There were significant differences and variation in the number of somatic mutation and DNA methylation events in the same gene in each type of breast cancer. A list of the top most highly somatic mutated and aberrantly methylated genes is presented in Table 3. A full list of somatic mutated genes, which were significantly differentially altered in the transcriptome (DE) and differentially altered in the methylome between TNBC and non-TNBC are presented in Supplementary Table S4. There were significant differences in somatic, DNA methylation and gene expression variation between TNBC and non-TNBC. The significant differences in DNA

methylation and somatic mutation profiles suggest that TNBC and non-TNBC may be amenable for somatic mutation-based and methylation-based classification, or a combination thereof.

**Table 3.** List of 30 somatic mutated genes significantly differentially expressed (DE), and also significantly differentially methylated (DM), between TNBC and non-TNBC.

Gene Symbol	Chromosome	DE <i>p</i> -Value	DM <i>p</i> -Value	DM Sites	TNBC Somatic Events	Non-TNBC Somatic Events
<i>RBM22</i>	5q33.1	$2.21 \times 10^{-10}$	$5.04 \times 10^{-10}$	7	3	
<i>CRIP1</i>	14q32.33	$2.75 \times 10^{-22}$	$4.17 \times 10^{-27}$	17	2	
<i>CYP51A1</i>	7q21.2	$8.95 \times 10^{-5}$	$1.39 \times 10^{-11}$	16	2	
<i>FAM227A</i>	22q13.1	$3.39 \times 10^{-3}$	$5.00 \times 10^{-4}$	5	2	
<i>FHL3</i>	1p34.3	$3.32 \times 10^{-5}$	$1.11 \times 10^{-15}$	9	2	
<i>GFI1</i>	1p22.1	$7.77 \times 10^{-5}$	$4.41 \times 10^{-30}$	63	2	
<i>HCP5</i>	6p21.33	$2.04 \times 10^{-3}$	$6.76 \times 10^{-10}$	39	2	
<i>HOXD13</i>	2q31.1	$6.69 \times 10^{-7}$	$5.98 \times 10^{-21}$	9	2	
<i>IGSF21</i>	1p36.13	$2.82 \times 10^{-7}$	$6.04 \times 10^{-28}$	50	2	
<i>P2RY2</i>	11q13.4	$3.33 \times 10^{-12}$	$1.59 \times 10^{-19}$	15	2	
<i>PCBP3</i>	21q22.3	$1.64 \times 10^{-8}$	$9.13 \times 10^{-24}$	42	2	
<i>RBM17</i>	10p15.1	$3.11 \times 10^{-15}$	$3.98 \times 10^{-8}$	6	2	
<i>RNF26</i>	11q23	$1.99 \times 10^{-3}$	$1.28 \times 10^{-4}$	4	2	
<i>SCAMP2</i>	15q24.1	$8.94 \times 10^{-9}$	$1.15 \times 10^{-9}$	4	2	
<i>SCNN1B</i>	16p12.2	$3.39 \times 10^{-2}$	$4.08 \times 10^{-17}$	9	2	
<i>MAP2K4</i>	17p12	$1.66 \times 10^{-8}$	$2.73 \times 10^{-12}$	7		33
<i>ZSWIM8</i>	10q22.2	$1.19 \times 10^{-4}$	$3.34 \times 10^{-9}$	4		18
<i>MALAT1</i>	11q13.1	$2.28 \times 10^{-11}$	$3.79 \times 10^{-5}$	4		16
<i>MED23</i>	6q23.2	$1.13 \times 10^{-3}$	$8.65 \times 10^{-6}$	9		16
<i>NRCAM</i>	7q31.1	$3.22 \times 10^{-4}$	$1.08 \times 10^{-29}$	17		16
<i>RAB3GAP2</i>	1q41	$3.04 \times 10^{-3}$	$9.49 \times 10^{-15}$	8		16
<i>KIAA0100</i>	17q11.2	$1.13 \times 10^{-6}$	$3.46 \times 10^{-7}$	5		15
<i>ABCC4</i>	13q32.1	$8.68 \times 10^{-8}$	$1.04 \times 10^{-18}$	17		13
<i>MED13L</i>	12q24.21	$1.46 \times 10^{-13}$	$7.41 \times 10^{-23}$	12		13
<i>RIPK1</i>	6p25.2	$1.87 \times 10^{-6}$	$3.25 \times 10^{-14}$	10		13
<i>TIAM2</i>	6q25.2-q25.3	$1.83 \times 10^{-4}$	$1.67 \times 10^{-23}$	36		13
<i>ADAMTS16</i>	5p15.32	$2.51 \times 10^{-4}$	$6.92 \times 10^{-17}$	32		12
<i>ADGB</i>	6q24.3	$5.90 \times 10^{-12}$	$1.05 \times 10^{-3}$	5		12
<i>PHF8</i>	Xp11.22	$7.01 \times 10^{-7}$	$4.42 \times 10^{-8}$	6		12
<i>RGS7</i>	1q43	$1.30 \times 10^{-4}$	$2.95 \times 10^{-18}$	26		12

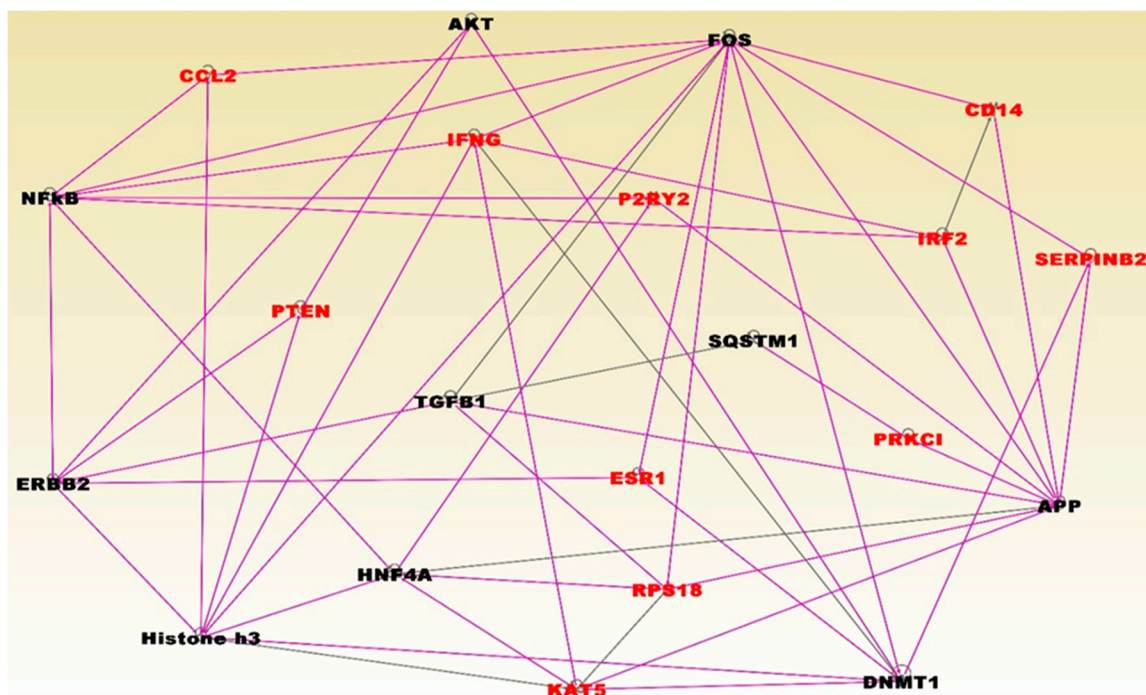
Further evaluation of the frequency of somatic mutations in each type of breast cancer revealed higher mutation occurrence in non-TNBC, compared to TNBC (Table 3). In addition, there was an uneven distribution of somatic mutation and epigenomic events. This suggests that somatic and epigenomic alterations are likely to occur at different times during the course of each disease. The results confirm our hypothesis that, for a selected set of genes, there were significant differences in somatic mutation burden, DNA methylation and gene expression levels between TNBC and non-TNBC (Table 3). As noted earlier in this report, the significant variation in somatic mutations, DNA methylation and gene expression levels can partially be explained by the heterogeneous nature of the two types of breast cancer.

### 3.5. Discovery of Gene Regulatory Networks and Biological Pathways

The development and progression of TNBC and non-TNBC involve both genomic and epigenomic alterations. Therefore, to address the hypothesis that somatic mutations and epigenetic alterations regulate and impact gene regulatory networks and biological pathways driving TNBC and non-TNBC, we conducted pathway prediction and network analysis, focusing on genes harboring somatic mutations and epigenetic alterations unique to each type of breast cancer. Here we sought to: (i) unravel the genomic-epigenomic interaction landscape in TNBC and non-TNBC; (ii) gain insights about the broader biological context in which genomic and epigenomic factors operate; and (iii) to establish putative

functional bridges between somatic mutations and epigenomic alterations and the pathways they affect. Such pathways could be potential therapeutic targets.

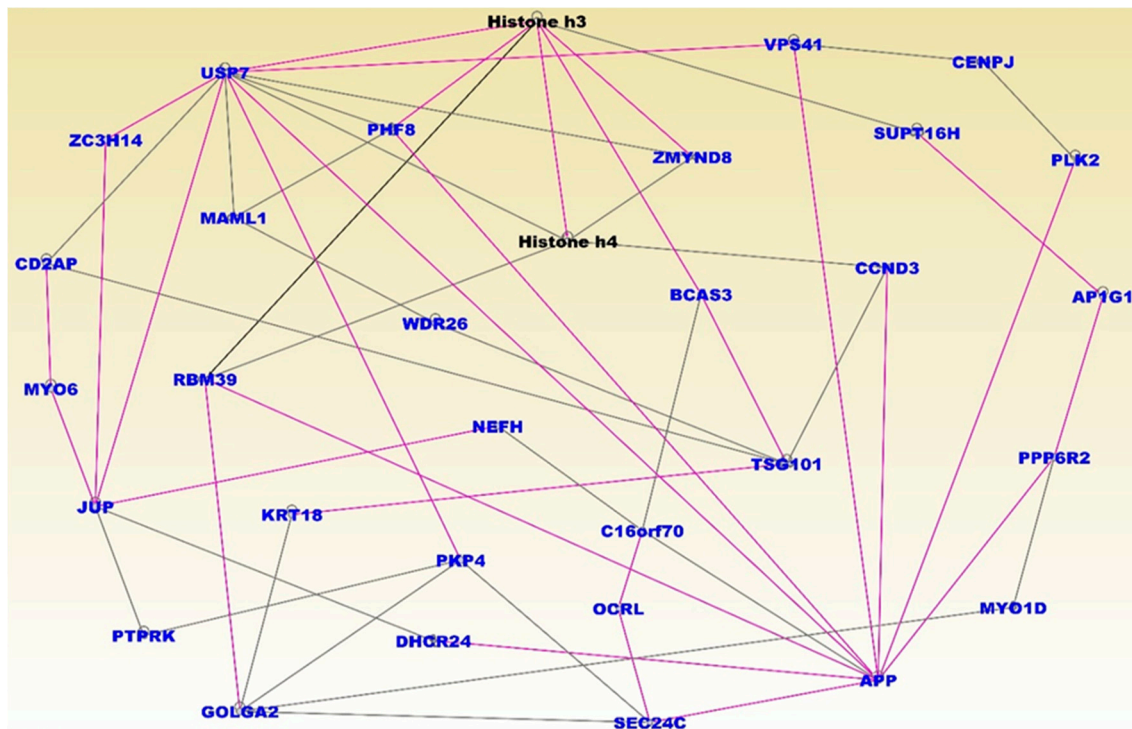
The results of network and pathway analysis for the top most significant networks for TNBC are presented in Figure 5. The highest scoring networks included the genes predicted to be involved in: hereditary diseases, organismal injury and abnormalities (Z-score = 35); cellular development, cellular growth, proliferation, protein synthesis (Z-score = 30); cell-to-cell signaling and interaction, immune cell trafficking, cellular development, cellular growth and proliferation, embryonic development (Z-Score = 28); cellular assembly and organization, cellular compromise, DNA replication, recombination, and repair (Z-Score = 25); and cell cycle, cellular assembly and organization, molecular transport (Z-Score = 21). Network analysis revealed functional relationships and interactions among aberrantly methylated genes containing somatic mutations in gene regulatory networks (see symbols in red fonts) for TNBC (Figure 5). Network analysis also revealed functional relationships and interactions with other genes (see black fonts) (Figure 5).



**Figure 5.** Molecular networks for genes enriched for somatic and epigenetic variation in TNBC. Genes are shown in nodes and vertices represent functional relationships. The gene symbols in red fonts represent genes containing somatic and epigenetic alterations in TNBC only. Genes in black font are other genes predicted to interact with genes containing somatic and epigenetic alterations. Note that, because the analysis involved merging the top networks containing genes with overlapping functions, the colors of the lines (vertices) are also overlapping to reflect this. Pink and black lines represent interactions among somatic and epigenetic altered genes (red fonts), and with predicted genes in black fonts.

The somatic mutated genes that were also aberrantly methylated and interacting in gene regulatory networks included the genes CCL2, PTEN, ESR1, IFNG, P2RY2, INF2, CD44, SERPINB2, PRKCI, RPB18 and KAT5 (Figure 5). Pathway analysis revealed important pathways implicated in TNBC. The top five most significant pathways enriched for somatic and epigenetic variation included the retinal biosynthesis, BAG2, LXR/RXR, EIF2 and P2Y purigenic receptor signaling pathways. The results of network and pathway analysis for the topmost significant networks for non-TNBC are presented in Figure 6. The highest scoring networks included the genes predicted to be involved in: cellular assembly and organization and cellular compromise (Z-Score 59); cancer, organismal injury and

abnormalities (Z-Score = 51); cell morphology and embryonic development, (Z-Score = 48); connective tissue disorders, organismal injury and abnormalities and organ development (Z-Score = 46); cellular compromise, cellular function and maintenance, cancer, (Z-Score 32); cancer and immunological disease (Z-Score = 28); cancer, cell death and survival, cellular development (Z-Score = 26). Network analysis revealed functional relationships and interactions among aberrantly methylated genes containing somatic mutations in gene regulatory networks (see symbols in red fonts) for non-TNBC (Figure 6). Network analysis also revealed functional relationships and interactions with other genes, including histone H3 and H4, which are epigenetically altered (see black fonts) (Figure 6).



**Figure 6.** Molecular networks enriched for somatic and epigenetic variation in non-TNBC. Genes are shown in nodes, and vertices represent functional relationships. The gene symbols in blue fonts represent genes containing somatic and epigenetic alterations in non-TNBC only. Note that the use of blue fonts for somatic and epigenomic altered genes here was designed to provide a contrast between the two types of breast cancer. Genes in black fonts are other genes predicted to interact with genes containing somatic and epigenetic alterations. Note that, because the analysis involved merging the top networks containing genes with overlapping functions, the colors of the lines (vertices) are also overlapping to reflect this. Pink and black lines represent interactions among somatic and epigenetic altered genes (blue fonts) and with predicted genes in black fonts.

The somatic mutated aberrantly methylated genes interacting in gene regulatory networks included the genes USP7, ZC3H14, CD2AP, MYO6, JUP, PTPRK, GOLGA2, PHF8, MAML1, RBM39, KRT18, PKP4, DHCR24, WDR26, NEFH, C16orf70, OCRL, SEC24C, APP, TSG101, BCAS3, CCND3, ZMYND8, VPS41, CENPJ, SUP16H, PLK2, AP1G1, PPP6R2 and MYO1D, which are represented in blue fonts (Figure 6). Pathway analysis revealed important signaling pathways implicated in non-TNBC. The most significant pathways included the UVB-Induced MAPK, PCP, Apelin endothelial, Endoplasmic reticulum stress and mechanisms of viral exit from host signaling pathways. Overall, integrative analysis revealed gene regulatory networks and signaling pathways enriched for somatic mutations and epigenomic alterations, unique to each type of breast cancer. The analysis also demonstrated that a subset of genes altered in the tumor genome, transcriptome and the methylome distinguished TNBC from non-TNBC. Both within type of breast cancer and comparison analysis of

gene expression and DNA methylation between the two types of breast cancer showed that DNA methylation affects gene expression in TNBC and non-TNBC, and that these effects differ between the two types of breast cancers, consistent with our previous report [22].

#### 4. Discussion

Since the completion of TCGA [19], large omics data sets have been made available. With the use of these data resources, we are now well-positioned to understand the molecular mechanisms underlying the etiological and biological differences and drivers of the two types of cancers. Here, we performed integrative data analysis combining multi-omics data for the discovery of integrated genomic, epigenomic and gene expression profiling signatures, networks and signaling pathways unique to TNBC and non-TNBC, and to discover signatures distinguishing the two types of breast cancer. The discovery of somatic mutated genes, which are also aberrantly methylated, suggests that genes driving tumorigenesis in TNBC and non-TNBC are under both genomic and epigenomic control. Here, we discuss the potential translational impact and clinical significance of the results of this investigation below.

##### 4.1. Novelty and Innovation of the Study

Our approach combined somatic, DNA methylation and gene expression data for the discovery of gene, somatic mutation and DNA methylation signatures unique to TNBC and non-TNBC and signatures, distinguishing the two types of breast cancer. Our approach represents a novel and unified approach to the discovery of potential clinically actionable biomarkers and targets for TNBC and non-TNBC, using multi-omics data. Discovery of integrated somatic mutation and DNA methylation signatures distinguishing the two types of breast cancer suggests that TNBC and non-TNBC may be amenable to mutation and DNA methylation-based classification. The discovery of signatures unique to TNBC and non-TNBC suggests that different biological mechanisms may be controlling the development and progression of each type of cancer. To our knowledge, this is the first study to combine three pieces of information (somatic, epigenetic and gene expression data) for the discovery of gene signatures unique to TNBC and non-TNBC, and signatures distinguishing the two types of breast cancer. Gene expression profiling to identify key differentially expressed genes and predict clinical outcomes and analysis of differences in serum protein mass spectrometry between TNBC and non-TNBC have been reported [35,36]. Breast cancer subtype identification using somatic mutations and DNA methylation have also been reported [37,38]. The novel and innovative aspect of this investigation is that it combined three types of omics data (somatic, epigenomic and gene expression variation) to identify signatures unique to TNBC and non-TNBC, and signatures distinguishing the two types of breast. This novel integrated genomics approach to discovery of potential clinically actionable molecular markers and potential targets for the development of novel therapeutics in TNBC and non-TNBC has not been previously reported.

##### 4.2. Complementing Omics Data with Standard Clinical Protocols

Traditionally, there have been several widely accepted methods for clinical classification of types of breast cancer. These include classification based on: (i) histology using morphological characteristics; (ii) immunohistochemical (IHC) markers, also used here to guide data analysis; and (iii) more recently, using gene expression profiles as the standard [8,10,35]. The significance of our investigation is that these traditional protocols could be complemented by the results in this investigation. Integrating somatic, epigenomic and gene expression data with traditional protocols holds promise not only for defining the types of breast cancer, but also for causally associating such types with the molecular mechanisms driving and distinguishing them. Moreover, integrating omics with standard protocols may help to eliminate errors associated with traditional protocols, such as IHC [8,10] and improve patient classification.

#### 4.3. Clinical Significance of Integrating Gene and Epigenomic Signatures in TNBC and Non-TNBC

Over the last decade, gene expression has gained prominence as the standard protocol in breast cancer [15–18]. Clinically validated prognostic gene signatures, such as PAM50 [15,16] and MammaPrint [17,18], are poised to improve clinical outcomes via precision medicine. One limitation and critical unmet need in the use of such assays is that they lack information on molecular drivers of the disease and biological factors regulating the tumor microenvironment. Given that somatic mutations drive tumorigenesis, enduring epigenetic landmarks define the tumor microenvironment [39], and DNA methylation affects gene expression [22], integrating information on gene expression, somatic mutation and DNA methylation, as demonstrated in this study, which provides an avenue for addressing this limitation. Although we did not explore the therapeutic potential of the discovered epigenomic markers in this study, several studies have reported the therapeutic potential and prognostic value of DNA methylation markers in breast cancer, specifically TNBC [40–43]. This is significant, because currently there are no effective targeted therapies for TNBC.

#### 4.4. Somatic Mutation Signatures in TNBC and Non-TNBC

The discovery of somatic mutation signatures unique to TNBC and non-TNBC and the differences in somatic mutation burden between the two types of breast cancer is of particular interest, because somatic driver mutations confer selective clonal growth advantage to tumor cells. Importantly, the discovery of a signature of somatic mutated genes distinguishing TNBC from non-TNBC suggests that the two types of breast cancer may be amenable to mutation-based classification. This could enable the identification of patients at high risk of developing aggressive tumors, such as TNBC, who could be prioritized for treatment. In this study, we did not investigate the mutation-based classification of the two subtypes of breast cancer. However, cancer subtype identification using somatic mutation data has been reported [37], though our study focuses on TNBC and non-TNBC, a problem which has not been addressed.

#### 4.5. Signaling Pathways as Potential Therapeutic Targets

Our investigation revealed biological pathways dysregulated by somatic mutations and aberrant DNA methylation, including retinal biosynthesis, BAG2, LXR/RXR signaling pathways in TNBC, which have the potential to serve as therapeutic targets if validated. The discovery of retinal biosynthesis signaling pathway is significant. This is because a major obstacle in the use of retinoid therapy in cancer treatment is tumor resistance to this agent [44]. One approach to overcoming the resistance of tumor cells to retinoic acid may be genomic and epigenomic reprogramming, targeting and suppressing this pathway using markers that have been shown to restore sensitivity to retinoic acid [44]. For example, curcumin has been shown to restore sensitivity to retinoic acid in TNBC cells [44], and to induce retinoic mediated apoptosis in retinoic acid resistant breast cancer cells [45]. Moreover, DNA methylation investigated here has been shown to predict the response of TNBC to all-trans retinoic acid [40]. This is particularly important, because retinoid signaling regulates breast cancer stem cell differentiation—a critical biological process in tumor development and progression [46].

Another pathway implicated in breast cancer discovered here was BAG2. The significance of this finding is that the Bcl-2-associated athanogene 2 (BAG2) has been shown to regulate the oncogenic functions of pro-cathepsin B in metastatic TNBC [47]. This makes the BAG2 discovered here a potential therapeutic target, which may reduce the burden of metastatic TNBC. We also discovered the LXR/RXR signaling pathway, which has been implicated in TNBC [48,49]. This is particularly interesting, because while our study was based on TCGA data, which is almost exclusively derived from women of European ancestry, the LXR/RXR signaling pathway has also been associated with TNBC in African American women using a different data set [49]. The significance of this finding is that African American women tend to have higher incidences and higher mortality rates from TNBC, and are affected at younger age—before the recommended age of 50 for screening [49].



We discovered signaling pathways associated with non-TNBC, notably the UVB-induced MAPK and apelin endothelial signaling pathways. The clinical significance of these findings is that the UVB-induced MAPK signaling pathway has been associated with estrogenic chemicals, most of which are considered carcinogenic [50]. Another pathway of clinical significance that we discovered in non-TNBC is the apelin signaling pathway [51–53]. This pathway was recently patented as a therapeutic target for breast cancer [51–53]. Taken together, these observations suggest that if confirmed, the discovered signaling pathways have therapeutic potential in TNBC and non-TNBC.

Apart from the discovered pathways, we also discovered a genomically and epigenomically altered ESR1 gene—a critical biomarker to both TNBC and non-TNBC. A negative ESR1, along with negative PR and lack amplification of HER-2, define TNBC. However, ESR1 also plays a major role in non-TNBC. Perhaps one of the most distinguishing features relevant to this investigation is that patients with estrogen receptor or hormone receptor-positive tumors (herein defined as non-TNBC) receive endocrine and targeted therapies [54], whereas TNBC is non-responsive, and there are currently no Food and Drug Administration [FDA] targeted therapies for TNBC. In a comprehensive review of the factors predictive of response to hormone therapy in breast cancer, Rastelli and Crispino confirmed the role of estrogen receptor content as a predictor of response to endocrine treatment [55]. They found that the benefits from endocrine treatment were directly proportional to estrogen receptor levels in non-TNBC [55]. Taken together, the discovery of a somatic mutated and epigenetic altered ESR1 in this study suggests that it could have potential therapeutic implications.

In sum, this investigation revealed that TNBC and non-TNBC are complex diseases, whose development and progression involve oncogenic interactions between somatic and epigenomic alterations, unique to each type of breast cancer. Our investigation further reveals that oncogenic interactions between these biological factors likely lead to the dysregulation of the gene regulatory networks and signaling pathways that drive TNBC and non-TNBC. We argue that integrating large-scale, high-dimensional somatic mutation, DNA methylation and gene expression data holds promise, not only for unraveling the genomic-epigenomic interaction landscape of TNBC and non-TNBC, but also provides the avenue for uncovering the molecular mechanisms characterizing the biological differences and the drivers of the two types of breast cancer.

**Limitations:** We have shown that integrative data analysis, combing information on somatic mutations, DNA methylation and gene expression, has the promise to unravel the genomic-epigenomic interaction landscape of TNBC and non-TNBC. However, the limitations of the study must be acknowledged. We used TCGA data, which is unique and optimal for integrative data analysis using disparate omics data from the same patient cohort. One key limitation worth mentioning here is that we did not use only driver somatic mutations in this investigation. Therefore, some of the somatic mutations used here may be passenger mutations, which do not contribute to TNBC or non-TNBC development. However, while passenger mutations may not contribute to disease development, they bear the imprints of the mutational mechanisms that have generated them by the process of natural selection. Therefore, they provide insights into the etiology and pathogenesis of TNBC and non-TNBC. Moreover, passenger mutations have been shown to accurately classify human tumors [56].

Another limitation worth noting here is that, in this study, we focused on TNBC and non-TNBC. Both types of breast cancer are inherently heterogeneous, each comprising of many subtypes not investigated here. Subtyping mapping was beyond the scope of this investigation, but is worth pursuing in future.

## 5. Conclusions

We discovered somatic, DNA methylation and gene expression signatures unique to TNBC and non-TNBC, and signatures distinguishing the two types of breast cancer. In addition, we discovered gene regulatory networks and signaling pathways enriched for somatic and epigenetic alterations in TNBC and non-TNBC. There were significant variations in patterns of mutation, DNA and gene expression profiles, reflecting the heterogeneity inherent in the two types of breast cancer.

The study demonstrates that integrative analysis of multi-omics data is a powerful approach to unravelling the genomic-epigenomic interaction landscape, and for the discovery of potential clinically actionable biomarkers, and targets for the development of therapeutic targets for TNBC and non-TNBC. Further research is recommended to map the genomic and epigenomic interactions in the subtypes of the two types of breast cancers to facilitate precision medicine.

**Data Sharing:** Availability of data and materials used in this study. The data that support the findings of this study are provided in supplementary tables as documented below, and original data sets are also made available in the TCGA, and are downloadable via the Genomics Data Commons <https://gdc.cancer.gov/>.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2072-6694/12/6/1559/s1>, **Table S1:** A complete list of somatic mutated genes and somatic mutations in TNBC and non-TNBC, **Table SE:** A complete list of genes significantly differentially expressed between TNBC and controls and between non-TNBC and controls along with their estimates of *p*-values, **Table SM:** A complete list of genes significantly differentially methylated between TNBC and controls and between non-TNBC and controls along with their estimates of *p*-values, **Table S2A:** A complete list of all the 12,816 significantly differentially expressed genes which were also differentially methylated in TNBC, **Table S2B:** A complete list of all the 13,460 significantly differentially expressed genes which were also differentially methylated in non-TNBC, **Table S2C:** A complete list of all the 10,365 genes significantly differentially expressed also significantly differentially methylated in TNBC and non-TNBC, **Table S3A:** A complete list of all the 4400 significantly differentially expressed genes which were also significantly differentially methylated distinguishing TNBC from non-TNBC, **Table S3B:** A full list of all the 5663 genes with binary pattern of presence, **Table S4:** A complete list of somatic mutated genes significantly differentially expressed (DE) which were also significantly differentially methylated between TNBC and non-TNBC.

**Author Contributions:** Conceptualization, C.H. and J.W.; methodology, C.H., J.W. and T.K.K.M.; software, J.W. and T.K.K.M.; validation, C.H. and J.W.; formal analysis, C.H., J.W., T.K.K.M.; investigation, C.H., J.W., T.K.K.M. and L.Z.; resources, C.H.; data curation, J.W. and T.K.K.M.; writing—original draft preparation, C.H., J.W., L.Z. and T.K.K.M.; writing—Review and editing, C.H., J.W. and L.Z.; visualization, C.H., J.W.; supervision, C.H.; project administration, C.H.; funding acquisition, C.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by internal funds from the LSUHSC-School of Medicine Startup funds and external funds from the UAB Center for Clinical Grant number # UL1TR001417 and The Louisiana Center for Translational Sciences LSUHSC # U54 GM12254691. All the views expressed in this manuscript are those of the authors and do not represent the funding sources or agency.

**Acknowledgments:** The authors wish to thank the participants who donated the samples to the TCGA project used to generate the data used in this project, technical support from TCGA and GDC staff as well as administrative staff from the Department of Genetics.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer Statistics, 2019. *CA Cancer J. Clin.* **2019**, *69*, 7–34. [[CrossRef](#)]
2. Dietze, E.C.; Sistrunk, C.; Miranda-Carboni, G.; O'Regan, R.; Seewaldt, V.L. Triple-negative breast cancer in African-American women: Disparities versus biology. *Nat. Rev. Cancer* **2015**, *15*, 248–254. [[CrossRef](#)]
3. Dietze, E.C.; Chavez, T.A.; Seewaldt, V.L. Obesity and Triple-Negative Breast Cancer: Disparities, Controversies, and Biology. *Am. J. Pathol.* **2018**, *188*, 280–290. [[CrossRef](#)]
4. Joyce, D.P.; Murphy, D.; Lowery, A.J.; Curran, C.; Barry, K.; Malone, C.; McLaughlin, R.; Kerin, M.J. Prospective comparison of outcome after treatment for triple-negative and non-triple-negative breast cancer. *Surgeon* **2017**, *15*, 272–277. [[CrossRef](#)] [[PubMed](#)]
5. Li, X.; Yang, J.; Peng, L.; Sahin, A.A.; Huo, L.; Ward, K.C.; O'Regan, R.; Torres, M.A.; Meisel, J.L. Triple-negative breast cancer has worse overall survival and cause-specific survival than non-triple-negative breast cancer. *Breast Cancer Res. Treat.* **2017**, *161*, 279–287. [[CrossRef](#)]
6. Pan, X.B.; Qu, S.; Jiang, Y.M.; Zhu, X.D. Triple Negative Breast Cancer versus Non-Triple Negative Breast Cancer Treated with Breast Conservation Surgery Followed by Radiotherapy: A Systematic Review and Meta-Analysis. *Breast Care* **2015**, *10*, 413–416. [[CrossRef](#)] [[PubMed](#)]

7. Ye, J.; Xia, X.; Dong, W.; Hao, H.; Meng, L.; Yang, Y.; Wang, R.; Lyu, Y.; Liu, Y. Cellular uptake mechanism and comparative evaluation of antineoplastic effects of paclitaxel-cholesterol lipid emulsion on triple-negative and non-triple-negative breast cancer cell lines. *Int. J. Nanomed.* **2016**, *11*, 4125–4140. [[CrossRef](#)] [[PubMed](#)]
8. Qiu, J.; Xue, X.; Hu, C.; Xu, H.; Kou, D.; Li, R.; Li, M. Comparison of Clinicopathological Features and Prognosis in Triple-Negative and Non-Triple Negative Breast Cancer. *J. Cancer* **2016**, *7*, 167–173. [[CrossRef](#)] [[PubMed](#)]
9. Podo, F.; Santoro, F.; di Leo, G.; Manoukian, S.; de Giacomi, C.; Corcione, S.; Cortesi, L.; Carbonaro, L.A.; Trimboli, R.M.; Cilotti, A.; et al. Triple-Negative versus Non-Triple-Negative Breast Cancers in High-Risk Women: Phenotype Features and Survival from the HIBCRI-1 MRI-Including Screening Study. *Clin. Cancer Res.* **2016**, *22*, 895–904. [[CrossRef](#)] [[PubMed](#)]
10. Nabi, M.G.; Ahangar, A.; Wahid, M.A.; Kuchay, S. Clinicopathological comparison of triple negative breast cancers with non-triple negative breast cancers in a hospital in North India. *Niger. J. Clin. Pract.* **2015**, *18*, 381–386. [[CrossRef](#)]
11. Koshy, N.; Quispe, D.; Shi, R.; Mansour, R.; Burton, G.V. Cisplatin-gemcitabine therapy in metastatic breast cancer: Improved outcome in triple negative breast cancer patients compared to non-triple negative patients. *Breast* **2010**, *19*, 246–248. [[CrossRef](#)] [[PubMed](#)]
12. Prat, A.; Pineda, E.; Adamo, B.; Galvan, P.; Fernandez, A.; Gaba, L.; Diez, M.; Viladot, M.; Arance, A.; Munoz, M. Clinical implications of the intrinsic molecular subtypes of breast cancer. *Breast* **2015**, *24*, S26–S35. [[CrossRef](#)]
13. Zhai, Q.; Li, H.; Sun, L.; Yuan, Y.; Wang, L. Identification of differentially expressed genes between triple and non-triple-negative breast cancer using bioinformatics analysis. *Breast Cancer* **2019**, *26*, 784–791. [[CrossRef](#)] [[PubMed](#)]
14. Darbeheshti, F.; Rezaei, N.; Amoli, M.M.; Mansoori, Y.; Bazzaz, J.T. Integrative analyses of triple negative dysregulated transcripts compared with non-triple negative tumors and their functional and molecular interactions. *J. Cell. Physiol.* **2019**, *234*, 22386–22399. [[CrossRef](#)]
15. Wallden, B.; Storhoff, J.; Nielsen, T.O.; Dowidar, N.; Schaper, C.; Ferree, S.; Liu, S.; Leung, S.; Geiss, G.K.; Snider, J.E.; et al. Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC Med. Genom.* **2015**, *8*, 54. [[CrossRef](#)]
16. Nielsen, T.; Wallden, B.; Schaper, C.; Ferree, S.; Liu, S.; Gao, D.; Barry, G.; Dowidar, N.; Maysuria, M.; Storhoff, J. Analytical validation of the PAM50-based Prosigna Breast Cancer Prognostic Gene Signature Assay and nCounter Analysis System using formalin-fixed paraffin-embedded breast tumor specimens. *BMC Cancer* **2014**, *14*, 177. [[CrossRef](#)]
17. Brandão, M.; Pondé, N.; Piccart-Gebhart, M. MammaPrint™: A comprehensive review. *Future Oncol.* **2019**, *15*, 207–224. [[CrossRef](#)]
18. Slodkowska, E.A.; Ross, J.S. MammaPrint 70-gene signature: Another milestone in personalized medical care for breast cancer patients. *Expert Rev. Mol. Diagn.* **2009**, *9*, 417–422. [[CrossRef](#)] [[PubMed](#)]
19. Weinstein, J.N.; Collisson, E.A.; Mills, G.B.; Shaw, K.R.M.; Ozenberger, B.A.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J.M.; Chang, K.; et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **2013**, *45*, 1113–1120. [[CrossRef](#)]
20. Shah, S.P.; Roth, A.; Goya, R.; Oloumi, A.; Ha, G.; Zhao, Y.; Turashvili, G.; Ding, J.; Tse, K.; Haffari, G.; et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **2012**, *486*, 395–399. [[CrossRef](#)]
21. Hudson, T.J.; Anderson, W.; Aretz, A.; Barker, A.D.; Bell, C.; Bernabé, R.R.; Bhan, M.K.; Calvo, F.; Eerola, I.; Gerhard, D.S.; et al. International network of cancer genome projects. *Nature* **2010**, *464*, 993–998. [[PubMed](#)]
22. Wu, J.; Mamidi, T.K.K.; Zhang, L.; Hicks, C. Deconvolution of the Genomic and Epigenomic Interaction Landscape of Triple-Negative Breast Cancer. *Cancers* **2019**, *11*, E1692. [[PubMed](#)]
23. The Genomics Data Commons. Available online: <https://portal.gdc.cancer.gov/> (accessed on 13 February 2020).
24. Marabita, F.; Almgren, M.; Lindholm, M.E.; Ruhrmann, S.; Fagerström-Billai, F.; Jagodic, M.; Sundberg, C.J.; Ekström, T.J.; Teschendorff, A.E.; Tegnér, J.; et al. An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform. *Epigenetics* **2013**, *8*, 333–346. [[CrossRef](#)] [[PubMed](#)]

25. Liu, J.; Siegmund, K.D. An evaluation of processing methods for HumanMethylation450 BeadChip data. *BMC Genom.* **2016**, *17*, 469. [CrossRef]
26. Maksimovic, J.; Gordon, L.; Oshlack, A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol.* **2012**, *13*, R44. [PubMed]
27. Wang, T.; Guan, W.; Lin, J.; Boutaoui, N.; Canino, G.; Luo, J.; Celedón, J.C.; Chen, W. A systematic study of normalization methods for Infinium 450K methylation data using whole-genome bisulfite sequencing data. *Epigenetics* **2015**, *10*, 662–669. [CrossRef]
28. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47. [CrossRef]
29. Ensemble Biomart Browser. Available online: <https://useast.ensembl.org/info/data/biomart/index.html> (accessed on 7 January 2020).
30. Abrams, Z.B.; Johnson, T.S.; Huang, K.; Payne, P.R.O.; Coombes, K. A protocol to evaluate RNA sequencing normalization methods. *BMC Bioinform.* **2019**, *20*, 679. [CrossRef]
31. Benjamini, Y.; Yosef, H. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **1995**, *57*, 289–300. [CrossRef]
32. TCGAvisualize-Starburst Tool. Available online: [https://rdrr.io/bioc/TCGAbiolinks/man/TCGAvisualize\\_starburst.html](https://rdrr.io/bioc/TCGAbiolinks/man/TCGAvisualize_starburst.html) (accessed on 7 September 2019).
33. Ingenuity Pathways Analysis (IPA) System Redwood, CA: Ingenuity Systems. Available online: <http://www.ingenuity.com/> (accessed on 7 September 2019).
34. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef]
35. Chen, B.; Tang, H.; Chen, X.; Zhang, G.; Wang, Y.; Xie, X.; Liao, N. Transcriptomic analyses identify key differentially expressed genes and clinical outcomes between triple-negative and non-triple-negative breast cancer. *Cancer Manag. Res.* **2018**, *11*, 179–190. [CrossRef] [PubMed]
36. Liu, A.N.; Sun, P.; Liu, J.N.; Yu, C.Y.; Qu, H.J.; Jiao, A.H.; Zhang, L.M. Analysis of the differences of serum protein mass spectrometry in patients with triple negative breast cancer and non-triple negative breast cancer. *Tumour Biol.* **2014**, *35*, 9751–9757. [CrossRef] [PubMed]
37. Kuijjer, M.L.; Paulson, J.N.; Salzman, P.; Ding, W.; Quackenbush, J. Cancer subtype identification using somatic mutation data. *Br. J. Cancer* **2018**, *118*, 1492–1501. [CrossRef] [PubMed]
38. Stefansson, O.A.; Moran, S.; Gomez, A.; Sayols, S.; Arribasjorba, C.; Sandoval, J.; Hilmarsdottir, H.; Olafsdottir, E.J.; Tryggvadottir, L.; Jonasson, J.G.; et al. A DNA methylation-based definition of biologically distinct breast cancer subtypes. *Mol. Oncol.* **2015**, *9*, 555–568. [CrossRef]
39. Pidsley, R.; Lawrence, M.G.; Zotenko, E.; Niranjan, B.; Statham, A.L.; Song, J.Z.; Chabanon, R.M.; Qu, W.; Wang, H.; Richards, M.; et al. Enduring epigenetic landmarks define the cancer microenvironment. *Genome Res.* **2018**, *28*, 625–638. [CrossRef] [PubMed]
40. Coyle, K.M.; Dean, C.A.; Thomas, M.L.; Vidovic, D.; Giacomantonio, C.A.; Helyer, L.; Marcato, P. DNA Methylation Predicts the Response of Triple-Negative Breast Cancers to All-Trans Retinoic Acid. *Cancers* **2018**, *10*, E397. [CrossRef]
41. Yu, J.; Qin, B.; Moyer, A.M.; Nowsheen, S.; Liu, T.; Qin, S.; Zhuang, Y.; Liu, D.; Lu, S.W.; Kalari, K.R.; et al. DNA methyltransferase expression in triple-negative breast cancer predicts sensitivity to decitabine. *J. Clin. Investig.* **2018**, *128*, 2376–2388. [CrossRef]
42. Szic, K.S.; Declerck, K.; Crans, R.A.; Diddens, J.; Scherf, D.; Gerhauser, C.; Berghe, W.V. Epigenetic silencing of triple negative breast cancer hallmarks by Withaferin A. *Oncotarget* **2017**, *8*, 40434–40453. [CrossRef]
43. Stirzaker, C.; Zotenko, E.; Song, J.Z.; Qu, W.; Nair, S.S.; Locke, W.J.; Stone, A.; Armstrong, N.J.; Robinson, M.D.; Dobrovic, A.; et al. Methylome sequencing in triple-negative breast cancer reveals distinct methylation clusters with prognostic value. *Nat. Commun.* **2015**, *6*, 5899. [CrossRef]
44. Thulasiraman, P.; McAndrews, D.J.; Mohiudddin, I.Q. Curcumin restores sensitivity to retinoic acid in triple negative breast cancer cells. *BMC Cancer* **2014**, *14*, 724. [CrossRef]
45. Thulasiraman, P.; Garriga, G.; Danthuluri, V.; McAndrews, D.J.; Mohiudddin, I.Q. Activation of the CRABP1/RAR pathway by curcumin induces retinoic acid mediated apoptosis in retinoic acid resistant breast cancer cells. *Oncol. Rep.* **2017**, *37*, 2007–2015. [CrossRef] [PubMed]

46. Ginestier, C.; Wicinski, J.; Cervera, N.; Monville, F.; Finetti, P.; Bertucci, F.; Wicha, M.S.; Birnbaum, D.; Charafejauffret, E. Retinoid signaling regulates breast cancer stem cell differentiation. *Cell Cycle* **2009**, *8*, 3297–3302. [[CrossRef](#)] [[PubMed](#)]
47. Yang, K.; Bae, E.; Ahn, S.G.; Pang, K.; Park, Y.W.; Park, J.; Lee, J.; Ooshima, A.; Park, B.; Kim, J.; et al. Co-chaperone BAG2 Determines the Pro-oncogenic Role of Cathepsin B in Triple-Negative Breast Cancer Cells. *Cell Rep.* **2017**, *21*, 2952–2964. [[CrossRef](#)] [[PubMed](#)]
48. Torresluquis, O.; Madden, K.; Ndri, N.M.; Berg, R.; Olopade, O.F.; Ngwa, W.; Abuidris, D.O.; Mittal, S.K.; Lyncook, B.D.; Mohammed, S.I. LXR/RXR pathway signaling associated with triple-negative breast cancer in African American women. *Breast Cancer* **2018**, *11*. [[CrossRef](#)]
49. Chen, X.; Wu, J.; Huang, H.; Ding, Q.; Liu, X.; Chen, L.; Zha, X.; Liang, M.; He, J.; Zhu, Q.; et al. Comparative Profiling of Triple-Negative Breast Carcinomas Tissue Glycoproteome by Sequential Purification of Glycoproteins and Stable Isotope Labeling. *Cell. Physiol. Biochem.* **2016**, *38*, 110–121. [[CrossRef](#)]
50. Kiyama, R.; Zhu, Y. DNA microarray-based gene expression profiling of estrogenic chemicals. *Cell. Mol. Life Sci.* **2014**, *71*, 2065–2082. [[CrossRef](#)]
51. Terasaka, S.; Inoue, A.; Tanji, M.; Kiyama, R. Expression profiling of estrogen-responsive genes in breast cancer cells treated with alkylphenols, chlorinated phenols, parabens, or bis- and benzoylphenols for evaluation of estrogenic activity. *Toxicol. Lett.* **2006**, *163*, 130–141. [[CrossRef](#)]
52. Wang, Z.; Greeley, G.H., Jr.; Qiu, S. Immunohistochemical localization of apelin in human normal breast and breast carcinoma. *J. Mol. Histol.* **2008**, *39*, 121–124. [[CrossRef](#)]
53. Rayalam, S.; Della-Fera, M.A.; Kasser, T.; Warren, W.; Baile, C.A. Emerging role of apelin as a therapeutic target in cancer: A patent review. *Recent Pat. Anticancer Drug Discov.* **2011**, *6*, 367–372. [[CrossRef](#)]
54. Waks, A.G.; Winer, E.P. Breast Cancer Treatment: A Review. *JAMA* **2019**, *321*, 288–300.
55. Rastelli, F.; Crispino, S. Factors Predictive of Response to Hormone Therapy in Breast Cancer. *Tumori* **2008**, *94*, 370–383. [[CrossRef](#)] [[PubMed](#)]
56. Salvadores, M.; Mas-Ponte, D.; Supek, F. Passenger mutations accurately classify human tumors. *PLoS Comput. Biol.* **2019**, *15*, e1006953. [[CrossRef](#)] [[PubMed](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).