ISSN: 2581-8651
Vol-3, Issue-6, Nov – Dec 2021
https://dx.doi.org/10.22161/jhed.3.6.9
Peer-Reviewed Journal

Journal of Humanities and Education Development (JHED)

Analyzing the influence of the mRNA levels and DNA copy number on Breast cancer-related protein expression.

Abdullah Othman Hassan[1], Rawaz Rizgar Hassan[2], Shang Ziyad Abdulaqadir[3], Rzgar Farooq Rashid[4]

[1,2,4]Department of Medical Laboratory Science, College of Science, Knowledge University, Kirkuk Road, 44001 Erbil, Kurdistan Region, Iraq
[3]Department of Biology, College of Science, Salahaddin University-Erbil, Iraq

Received: 04 Nov 2021; Received in revised form: 08 Dec 2021; Accepted: 16 Dec 2021

Abstract

Each protein's expression in breast cancers was examined to see if it was regulated at a specific molecular level, and whether this regulation was reflected in different molecular subtypes. Microarray technologies were used to examine the DNA, mRNA, and protein lysates of 251 breast carcinoma samples. It was possible to obtain data from all three levels of the PI3K/Akt pathway for 52 proteins that have been linked to cancer. In cis Spearman rank correlations between the three molecular levels were determined for every protein across all samples and for each intrinsic gene expression subtype, resulting in 63 comparisons total due to numerous gene probes matching to single proteins. The variance of subtype-specific correlation and the variations between overall and average subtype-specific correlation were used to study subtype-specific interactions between the three molecular levels. An external data collection of 703 breast tumor specimens was used to confirm the findings. There were four categories of proteins based on rank correlation values between the three molecular levels, with the proteins being sorted into each category. There were eight proteins in Group A that had a significant association between DNA copy number and mRNA expression (Bonferroni adjusted p 0.05), as well as the other way around. There were 14 proteins in Group B with a strong connection between mRNA levels and protein levels. There was a strong link between copy number and mRNA expression in the 15 proteins in Group C. No relationships were found between the remaining 25 proteins (group D). Identifying favorable associations between copy number levels, mRNA and protein expression was possible only when cancers were categorized according to their intrinsic subtype. Particular attention was paid to protein pairings that indicated large variance in correlation values between subtypes or the overall dataset. The basal-like subtype had the highest protein expression of cleaved caspase 7 and the lowest levels of hsa-miR-29c, which coincided with CASP7 gene expression. Low expression of CASP7 cleavage and low association to CASP7 gene expression were found in the luminal A-like subtype. This miRNA was found to have an apparent target sequence in CASP7 mRNA. As such, the hsa-miR 29c miRNA may be acting as a suppressor of CASP7 translation inside the luminal-A subtype. CCNB1 copy number and gene expression were shown to be unrelated across the entire cohort. Cyclin B1 mRNA and protein expression were found to be favorably associated to copy number data in most gene intrinsic subtypes, showing that copy number can influence overall protein expression. Each subtype's patients with decreased overall survival were recognized by aberrations of cyclin B1 copy number, as well. Genes and their products could be grouped into four groups based on the association between the three molecular levels, and the expression was anticipated to be regulated at distinct levels. Subtype-specific regulation was discovered after further stratification of the data set.

Keywords— ***Breast cancer, DNA copy, mRNA levels.***

I. INTRODUCTION

There are many different types of breast cancer, both in terms of the molecular and clinical aspects of the illness. Sub-groups of breast tumors have been identified based on features at the DNA, mRNA, and protein levels (referred to as "molecular levels"). Immunohistochemical staining methods are commonly used to evaluate the protein expression of estrogen receptor (ER), progesterone receptor

(PR), and human epidermal growth factor receptor (HER2/ERBB2), which are used in prognosis evaluation and therapy decision-making (Bareche, , et al., 2018). To enhance outcomes for breast cancer patients who express these proteins, the use of therapies targeting ER and HER2 (Geyer et al., 2006; Murphy and Fornier 2010; Early Breast Cancer Trialists' Collaborative Group, 1998) has been shown to be beneficial. As well as an increase in the level of ERBB2, there is a strong correlation between increased levels of HER2, as well as an increase in the number of copies of ERBB2 on chromosome. (Lin, , et al., 2018). While ER protein expression and ESR1 gene copy number gain are closely linked, copy number growth in ESR1 is rare (Basudan, et al., 2019). It is clear that these two receptors have different regulatory mechanisms, as the low frequency of ESR1 chromosomal abnormalities cannot explain the wide range of gene and protein expression (ESR1 and ER, respectively). mRNA and protein abundance is controlled by a variety of processes, including transcription factors, miRNA, linc- RNA, DNA methylation, translational and post-translational modifications, and protein stability (Johnstone, , et al., 2018). In signal transduction pathways such as the PI3K/Akt system, protein activity and protein communication are controlled by the phosphorylation and dephosphorylation of certain protein epitopes. Proliferation and cell death, as well as homeostasis, synaptic signaling and glucose metabolism are all controlled by the phosphorylation of the Akt proteins on the serine and threonine residues of many proteins. Activation and deregulation of the PI3K/Akt pathway are typical during carcinogenesis, in contrast to when cells are in a state of quiescence. Cancer biology and therapy discovery depend on a thorough understanding of the molecular mechanisms involved in deregulation of this system (Castaneda et al., 2010). In order to find out if the link between copy number levels and mRNA expression differed between the previously discovered molecular intrinsic subtypes of breast cancer, we set out to conduct a study (Shekha, et al., 2013). The PI3K/Akt pathway and other cancer-related processes were studied using 52 proteins known to be involved in this study. We used aCGH, whole genome gene expression microarrays, and RPPA to examine the DNA copy number, mRNA expression, and protein expression of 251 breast cancers. In order to categorize the 63 possible pairings of 52 distinct proteins into four groups, correlation studies were used.

II. MATERIALS AND METHODS

Comparative Genomic Hybridization (aCGH e array). Comparative Genomic Hybridization 244K Agilent Microarrays were used to measure DNA copy number. Oligonucleotide probes (60-mers) encompassing both coding and non-coding genomic areas are included in the array. There is an average of 7.4 kb of space between coding and non-coding regions (Xiao, , et al., 2018). In the past, we've discussed how to handle experimental data (Baumbusch et al., 2008; Russnes et al., 2010). Out of 251 patients, 196 had aCGH data.

The Human Genome Survey Microarray version 2.0 was used to evaluate gene expression (Applied Biosystem). Coverage of 29,098 transcripts is provided by a total of 32,878 60-mer probes on the genome array. In all, 194 unique samples with high-quality gene expression data were obtained after QC-parameters of the internal controls were used to determine the quality of the arrays. As a result, all genes coding for selected proteins were studied without any screening of probes. Quantile-normalized, log2-transformed and median gene-centered data were obtained. 194 of the 251 patients had gene expression data available. mRNA expression data may be found in GEO (GSE24117), and the experimental details of this dataset have already been disclosed (Buccitelli & Selbach, 2020).

The one-color microarray "Human miRNA Microarray Kit (V2) (Agilent Technologies) was used to measure miRNA expression in accordance with the manufacturer's procedure (miRNA Micro- array System v1.5). Based on miRBase release 10.1, this miRNA array includes 723 human and 76 viral miR-NAs. For each miRNA, the array has 15,744 features (60-mers) and 953 control probes, which means that each miRNA is replicated w18 times on it. For labeling and array hybridization, 100 ng of total RNA was utilized as input. Total RNA was extracted using TRIzol (Invitrogen). Agilent Scanner G2565A was used for scanning. Feature Extraction (FE) version 11, log2-transformed and quantile normalized in Genespring v.10.0 were used to process miRNA data (Agilent Technologies). FE's QC parameters were used to evaluate the product's quality. A total of 408 miRNAs were found to be expressed in this collection of human breast tumors after excluding those lacking a call in more than 80% of the samples.

The aCGH- and mRNA probes were matched to the corresponding 52 proteins from the RPPA platform in order to undertake cis-correlation (correlation between DNA, mRNA, and protein for the same genomic region) analyses across the three molecular levels. Data from each antibody was matched to the same gene expression probe for proteins when expression data were available for both total protein and phosphorylated epitopes (e.g. ER, ERp118, ERp167 were all matched to ESR1 gene probe). Expression of a multi-compartment protein was affected in cases when more than one mRNA transcript translated for the same protein subunit.

Adjuvant treatment was administered to all of the patients. This regimen was used for premenopausal women who had not yet begun their menopause treatment, and it included the administration of CMF (cyclophosphamide, methotrexate, and fluorouracil) over the course of eight or nine cycles, depending on whether they received radiation therapy (Rashid & Saler, 2018). For one year, postmenopausal women were administered Tamoxifen 30 mg daily as part of the "DBCG c" regimen. Patients were monitored throughout the first decade of their recovery, or until they experienced their first recurrence, death, or development of a new primary malignancy. recurrence in the same region (LRR), metastases to other parts of the body (DM), contralateral breast cancer (CBC), and death were the study's primary outcomes (Naidoo, , et al., 2018). A total of four additional patient cohorts, dubbed "MDG," "Uppsala," "ULL," and "MicMa," were added to the original dataset for CCNB1 copy number survival studies (Hasna, , et al., 2018). A total of 703 invasive breast carcinomas were included in the The Cancer Genome Atlas data portal (TCGA; http://tcga-data.nci.nih.gov/tcga/ (Rashid, 2021) to evaluate the correlations. Dry ice was used to freeze the tumor specimens, which were then sliced into three pieces and analyzed in triplicate. Total RNA, DNA, and proteins were extracted from individual parts of the central piece. Hematoxylin and eosin (HE) staining was performed on sections produced from the two adjacent portions. DNA copy numbers, mRNA expression, and protein expression data were generated from a total of 251 tumor tissues.

Gene Expression Subtyping

According to Sorlie et alintrinsic .'s gene list, tumor samples were subtyped using gene expression data (Guha, , et al., 2018). We found 374 gene probes on our gene expression platform by matching entrez gene id and gene symbol from the original intrinsic gene list, which had 561 gene probes. For these genes, the median-centered gene expression data was collected from 194 samples. The 374 gene probes were used to assess Pearson correlations between each tumor specimen and each of the five subtypes of centroids. Centroid-specific subtypes were assigned to each sample based on its strongest correlation. A Pearson correlation threshold of at least 0.15 was used in this experiment. "Unclassi-fied" samples had a maximum correlation of less than 0.15. As a result, Luminal A (n 1–63), Luminal B (n–25), HER2-enriched (n–27), Basal-like (n–29), and Normal-like (n–32; n–18 unclassified) were the most common classifications.

Inter-platform correlation between genes and proteins

The KolmogoroveSmirnov (KS) test was used to determine if each dataset had a normal distribution. This test calculates the maximum distance between the cumulative frequency curve and the best-fit normal curve of the data and assesses the statistical significance of this distance. KS-test As long as the p-value for the KS test is below 0.05, the normal distribution hypothesis is invalidated. The normalcy assumption is not disproved if the KS test has a high p-value. Each genomic DNA probe, mRNA expression probe, and antibody was tested for normalcy in SPSS (protein expression). Only the copy number data yielded extremely significant p-values, while the expression data yielded non-significant p-values. Spearman rank correlation was used in all comparisons because of the non-normal distribution of aCGH data (Pala, et al, 2021). Using SPSS 17, the Spearman rank cis-correlation was estimated on all three platforms for each pair of geneegenes or geneeproteins. Statistically significant positive correlations (p-value 0.05) were found after Bonfer- roni adjustment when Spearman correlation values were >0.3. Values were classified as high (>0.5), medium (0.5-0.3), low (0.3-0.15), none (0.15-0.15), and negative (0.15) depending on their level of correlation. Using 154 breast tumor samples, correlations between copy number and mRNA expression were evaluated for the selected genes. 164 breast tumor samples were used to find correlations between mRNA expression and protein expression, as well as correlations between copy Protein arrays in reverse phase

In order to examine the expression of individual proteins across several samples simultaneously, a reverse phase protein array (RPPA) was used (Tibes et al., 2006). An antibody specific for a protein of interest is used to incubate each array in the reverse phase array format, which immobilizes the samples to be examined on a nitrocellulose-coated slide (Jung, et al., 2019). PI3K/Akt pathway proteins and others associated with cancer were the primary targets of the antibodies used in this study. A log2-transformed protein-centered version of the RPPA data was used. For 210 of the 251 tumors, RPPA data was available. previously described and published RPPA dataset (Hennessy et al., 2010).

Each gene probe was paired with a specific protein (e.g. AMPK protein consists of gene products from seven different genes PRKAA1, PRKAA2, PRKAB1, PRKAB2, PRKAG1, PRKAG2, and PRKAG3). Gene probes for the same gene (i.e., a gene product) were matched to the appropriate protein in cases where numerous gene probes were available for the gene (total protein and phosphorylated epitopes). Probe IDs "186013" and "15991" were both found to be associated with IGFR1p and IGF1R. Finally, all of the aCGH probes in the gene's genomic area were matched to the gene expression probes. A single gene's genomic area was represented by the same number of copies in each sample. The signal from a single probe was therefore used as a proxy for the entire genomic region in question.

There were cases where gene expression and aCGH platforms disagreed, and the aCGH probe with the greatest correlation value was chosen. Supplementary Table 1 provides all of the matched information.

There were 135 potential combinations of protein epitopes (including phosphorylated epitopes), acgh probes, and gene probes for 52 distinct proteins. One DNA and gene-probe for each unique gene/protein representation for each molecular level (DNA, mRNA, and protein) was selected to reduce redundancy in the comparison, choosing probes that provided the highest cis-correlation for each pairing. The complete protein antibodies were selected for most pairings, except for HER2 and Stat6, because of technical difficulties with these two total-protein directed antibodies. In this case, the phosphorylated epitopes HER2-p1248 and Stat6-p641 were used instead of the entire protein for these proteins. As the activated form of the protein, we could analyze its functional form, HER2-p1248. RPPA data was only available for antibodies targeting the cleaved (and active) form of the proteins caspase 7 and PARP. All genes were included in proteins that had several gene products. Phosphorylated epitope-to-total protein and copy number-to-gene expression correlation comparisons can be found in the Supplementary material section (Supplementary Table 2). The supplemental data files 1, 2 and 3 contain the copy number logR-values, mRNA expression and protein expression for the pairings. A total of 161 breast cancers were used to identify the number and protein expression of 62 gene/protein pairs (52 distinct proteins).

Correlation within five gene expression subtypes

An analysis was conducted to examine the relationship between the three levels of expression and each molecular subtype. For a total of 140 samples, the in cis-correlations between copy number and mRNA expression were computed. Correlations between the expression of mRNA and the expression of protein in 151 samples from five different types were calculated. Samples were split into five sub-types based on copy number and protein expression, and the correlations were determined for 122 samples (Table 1).

The variance of the subtype-specific correlations was measured in order to evaluate subtype-specific correlations (the variance of these correlations were called VSC). High VSC indicated probable variations in gene and protein regulation among subtypes.

A variable called "Z" was created in order to compare the overall correlation to the correlations between individual subtypes. Assume "Ai" is the protein's overall Spearman correlation coefficient (Ai) and "Bi" is the weighted mean (weighted by the number of samples in each subtype) of the five subtype specific Spearman correlation coefficients (B). Positive Z indicates lower subtype-specific correlations compared to protein's overall correlation and negative shows higher subtype-specific correlations compared to the overall correlation for that protein.

Table 1: Sample of molecular levels

Differences	Normal	Basal	HER2-enr	Luminal A	Luminal B	Total
CN-PX	19	18	17	39	21	114
GX-PX	22	21	19	49	24	135
CN-GX	22	26	19	46	19	132

Targets-can detected all the miRNAs expected to have "conserved" targets in the base sequences of PECAM1 (CD31) and CASP7's two genes (proteins) (Bulut, et al., 2020). Five miRNAs with predicted targets in CASP7 and two miRNAs with predicted targets in PECAM1 were found after filtering the miRNA data by present call in at least 20% of the samples. Between the five expression subtypes, ANOVA was used to determine whether there were significant differences in protein and miRNA expression (cleaved caspase 7 and CD31 and targeted miRNAs).

The samples were divided into three groups based on cyclin B1 copy number and expression levels (GAIN/HIGH, NONE/MEDIUM, and LOSS/LOW) for each molecular level. A larger dataset of 506 cancer patients with an average total survival time of 15 years was used to boost the number of gains and losses at the copy number level. A copy number value of (LogR) >0.20 was assigned to the GAIN group, a value of (LogR) 0.20 was assigned to the LOSS group, and the rest were assigned to the NONE group.. The samples were divided into groups based on the 25th and 75th percentiles of expression for mRNA and protein. KaplaneMeier plots and log rank tests were used to evaluate the significance of overall survival for CCNB1/cyclin B1 expression in the CN, GX, and PX tissues using a 15-year overall survival as a result.

We used an external set of validation data to verify the accuracy of our correlation coefficients. The Cancer Genome Atlas database (Tanioka, , et al., 2018) had information on 39 of the 52 protein pairs for a total of 703 invasive breast cancers, including information on copy number, mRNA expression, and RPPA. Spearman rank correlations were calculated for the three comparisons and the 39 protein pairs with a minimum of 388 samples overlapped. There were 39 protein pairings in total, and the correlation values from the validation set and the origin- al set were compared to see if there was a comparable pattern of correlation (PAM50) (PAM50). A total of 220 luminal A and 127 luminal B were studied, as well as 55 luminal A and 93 luminal B, as well as 8 luminal A and 8 luminal B. This link was found in all 503 samples as well as within each category.

III. RESULTS

DNA copy number, mRNA expression, and protein expression all had their in-cis correlations computed for the 52 proteins tested (63 pairings to mRNA probes). To categorize the proteins into four groups, AeD-plots scatterplots between the three levels of each protein from each of the four groups, based on correlation values Copy number was strongly associated to mRNA expression (Spearman's rank correlation >0.3) for eight pairs in Group A, as were mRNA expression and protein expression. Five of the proteins (HER2p1248, p70S6 Kinase, IGFR1, Rab25, and PDK1) had a substantial correlation with protein expression, whereas the remaining three had a positive correlation (but not significantly) with copy number (4ebp1, Cyclin D1 and stat3). The fact that the expression of these proteins was positively correlated at all three molecular levels suggested that copy number differences influenced their expression. jlogRj > 0.4 was found in more than 15% of the samples for six of the eight proteins studied, indicating high copy number growth or loss. All other genes, IGF1R and STAT3, saw gains of 5% and 7%, respectively, in copy number. As compared to groups B, C, and D, group A had an average of four-fold more variation in copy number values, which supports copy number as a primary driver of protein expression. No significant association was seen between mRNA expression and protein expression in the 14 protein pairs in group B. In addition to these proteins, cKit and ER, Caveolin 1 and PR, cyclin B1 and Cyclin E1 were also included. Akt and p110a were also included. Although there were 14 proteins in group B, only five showed jlogRj > 0.4 in more than 15% of samples in comparison to group A. To put it another way, the lesser variability in copy number levels reduced the possibility that copy number abnormalities of these genes would be key drivers of mRNA and protein production. Similar to what has been shown with PTEN, Rb, BRCA1, MEK1, and p53. Only the copy number and mRNA expression had a significant association (Spearman's rank correlation > 0.3) for the 15 pairings in Group C. Here, we identify cleaved PARP, S6, TSC2, b-catenin, MAPKs and BRCA1, MEK1, E-cadherin, mTor, ERK4 and AMPK (for AMPK significant correlation was found for one out of 7 pairings to genes encoding subunits of this protein). Group C showed jlogRj > 0.4 in over 15% of samples in just 4 out of 15 genes, but group B showed no link between copy number and jlogRj > 0.4. As a result, the poor association between copy number and mRNA expression in group B was not only due to low copy number variance. Group D: mRNA and copy number levels have no meaningful association with protein expression. Group D was comprised of the remaining 25 couples whose molecular levels revealed no significant link. It comprised members such as p53, Bcl2, AMPK (6 genes that code for 6 of the protein complex's 7 subunits), p27, GSK3 (two genes that code for two subunits), XIAP, Stat6p641, LKB1, cJun, VEGFR2, Src, p21, Collagen VI (Sthmin), EGF receptor 2, EGFR receptor 3, JNK receptor 4, P38 and SGK jlogRj > 0.4 was found in just two of the 25 genomic areas in group D, which is less than 15% of the samples. It is possible that this group's lack of copy number-driven mRNA and/or protein expression could be explained by this low number of copy number abnormalities. Due to either a real lack of association or inadequate variety in copy number levels, the correlation between copy number and mRNA/protein expression may be low. Samples from the top and bottom 10% of copy number values were picked for each protein pairing in order to reduce the dilutive influence of non-aberrant cases (data not shown). As a result, five more proteins were identified for which copy number levels were found to correlate strongly with protein expression: Rb (RB1), (TP53), BRCA1 (BRCA1), MEK1 (MAP2K1), and PTEN (PTEN) (PTEN). The correlation between copy number and mRNA expression of Akt, p110a, cMyc, EGFR, CD31, and PAI1 was also observed (data not shown). Because of the lack of copy number variability, this shows that correlation can be reduced. Within each intrinsic molecular subtype, correlations were generated to investigate particular regulatory interactions (Sun, , et al., 2019). These within-subtype correlations' variance (hence referred to as variance) was then calculated. It was calculated for each protein to calculate the variance of subtype-specific correlations (VSC) for each comparison. A high VSC indicates protein pairs that reveal considerable differences in molecular correlations. Variable levels of variability in expression values at one or more molecular levels may be to blame for these discrepancies across the subtypes, or they may indicate variances in molecular regulatory relationships. For subtype-specific miRNA regulation, CASP7 and PECAM1 were chosen because of their high VSC. An external dataset from the Cancer Genome Atlas (TCGA) (http://tcga-data.nci.nih.gov/) was used to verify the subtype-specific correlation (see Materials and Methods). For most subtypes, positive and significant correlations were found between the original dataset's correlation values and the validation set's. According to comparisons of copy number and protein expression levels, the VSC of cleaved caspase 7 protein was higher than expected. Correlations between subtypes ranged from 0.27 (normal-like) to 0.65 (exceptional) (HER2 enriched). Using multigroup comparisons at the three molecular levels, we investigated whether these correlation discrepancies may be attributed to changes in CASP7 expression between subtypes. A KruskaleWallis test found a significant difference in CASP7 copy number between the HER2-enriched and normal-like subtypes (p 14 0.04), but ANOVA analyses of mRNA and protein expression found no significant differences in mRNA expression levels of CASP7 between any of the subtypes (the most significant p-value was 14 0.07 between luminal B and Her2-enriched). Cleaved caspase 7 protein expression was significantly higher in the basal-like subtype compared to all other subtypes (p 0.001). After all, cleaved caspase 7 is targeted by the antibody, which suggests that total protein level contributes to cleavage at least in part. CASP7 mRNA expression could not explain the differences in protein expression, so the expression of six miRNAs expected to target CASP7 mRNA (www.tar-getscan.org) was investigated (supplementary data file 4). As a theory, it was posited that The increased expression of miRNAs targeting CASP7 may have resulted in a corresponding inhibition of translation. There were three miRNAs that were projected to target CASP7 mRNA; hsa-miR-29c, hsa-miR-29c* and hsa-miR-29c. The pattern of hsa-miR-29c (and hsa-miR-29c*) expression was strikingly similar to that of cleaved caspase 7 protein expression in each of the five subtypes studied. It's possible that reduced hsa-miR-29c translational regulation explains the increased degree of cleaved caspase 7 expression in basal tumors. Each protein had a Z parameter calculated for each of the three correlation comparisons. A discrepancy between the "overall correlation" and the average of the five subtype-specific correlation values, weighted by the number of samples in each subtype, is described by the coefficient Z. As a result, Z measures the disparity between across-subtype correlation and the average correlation within each subtype. Low correlations within subtypes, combined with strong inter-subtype variability, resulted in a high total correlation. This suggests that significant correlations within subtypes were obscured by intersubtype differences, which resulted in a low overall correlation. A positive Z score for HER2p1248 (ERBB2) in the copy number to mRNA expression and mRNA expression to protein expression comparisons, and the second biggest score (after PDK1) in the copy number to protein expression comparison, was obtained. However, a more homogeneous ERBB2 expression in all but luminal B may account for the lower subtype-specific correlation values, as demonstrated in Supplementary. In both the copy number to mRNA expression and the copy number to protein expression comparisons, the greatest negative Z was identified for cyclin B1 (CCNB1). Having a negative Z indicates that the overall correlation (over the entire dataset) is lower than the mean correlation value for this pairing across all five subtypes. Because supervised subtyping minimizes the within-group variance and hence penalizes situations that achieve a positive correlation, a negative Z is an uncommon phenomenon that can reveal a hidden regulatory link that is being obscured by the overall dataset. For CCNB1 copy number, there was an overall correlation of 0.07, whereas the weighted mean subtype correlation of 0.33 (Z 14 0.40) was found. Although there was no overall link between CCNB1 copy number and mRNA expression, a positive correlation was detected in each of the five subtypes (significant p 0.05 in four subtypes) Breast cancer patients with elevated levels of cyclin B1 protein have been linked to a poor prognosis (Zhao, , et al., 2019). Both higher mRNA and protein expression were associated with a considerably worse 15-year overall survival rate, in agreement with previous investigations. Because the number of samples with CCNB1 copy number aberrations was minimal, this influence on overall survival was subsequently evaluated in an enlarged tumor collection that included copy number data (aCGH) from 506 breast cancers (Supplementary datafile 5). Because mRNA and protein expression data were not available for this extended sample set, sub-types could not be formed for all samples.. Because of this, ER-status stratification was used to create groups with enough incidents. When CCNB1 copy numbers were lost (LOSS-group), overall survival was lowered (p-value less than 0.001). Overall survival was also lowered in the GAIN-group (p-value: 0.068) compared to the NONE group (p-value: 0.068). Breast cancer subtypes differed in terms of cyclin B1 levels at all three molecular levels. With regard to CCNB1 copy number, basal-like subtypes were related with copy number loss, while luminal subtypes (associated with ER positive) were associated with copy number gains. The CCNB1 GAIN group had a poorer overall survival rate than the NONE group in patients with ER-positive malignancies, although the difference was not statistically significant (p-value 14 0.112). (no tumors in the ER-positive group showed loss of CCNB1). ER-negative and/or basal-like tumors may have contributed to the LOSS-decreased group's overall survival. Because of copy number induced higher expression of cyclin B1 in the GAIN-group, overall survival was lowered.

Discussion

Transcription, translation, and protein stability all play a role in protein levels. Variations in the amount of protein abundance variance explained by mRNA abundance are observed in the estimates (Rashid, 2017). It is clear that mRNA expression correlates significantly with protein production for around 35% of the proteins in our panel. Since mRNA expression data cannot account for all data on protein expression (despite being statistically significant), these correlations are not perfect. Because the degree of chromosomal aberrations in a particular genomic area has an effect on copy number levels, it has been proposed that copy number variation accounts for about 18% of the diversity in gene expression observed across disorders (Schulz, , et al., 2018). A variety of subtypes of breast cancer, as well as individual tumors, exhibit varying degrees of copy number aberrations (Rashid & Basusta, 2021). Consequently, it is likely that the number of genes and proteins whose expression is affected directly by copy number abnormalities varies between subtypes. More than a third of the genome's genes (35%) were shown to be affected by copy number, as well as around 12e15 percent of the proteins studied. The study, on the other hand, was based on a subset of proteins chosen for their potential roles in the development of breast cancer rather than to represent the entire proteome. Antibodies directed for a protein's functionally active configuration were also used to test several other proteins (eg. HER2p1248, Cleaved caspase 7, Cleaved PARP). Because of this, it is not possible to extrapolate the percentages of significantly associated pairs to the entire genome or proteome.

Correlation Test

Despite the fact that correlation does not necessarily imply causation, the biological relationship between DNA copy number levels, mRNA, and protein expression makes the assumption of causation less bold. A genomic region must be affected by aberrations in order for its copy number levels to be recognized as a likely driver of altered mRNA and protein expression. Copy number abnormalities have been found in the majority of genomic regions corresponding to proteins in group A in breast cancer (Craze, , et al., 2018). As a result, it was hypothesized that the eight proteins in group A were significantly influenced by copy number .

All of the proteins in group B had low or no association with copy number, however the mRNA expression to protein expression correlation was moderate ($R > 0.3$) or strong ($R > 0.5$). As a result, mRNA expression levels (in cis) but not copy number were expected to influence protein expression in group B. It is likely that these proteins are regulated by mechanisms other than copy number abnormalities (e.g. transcription factor activity, methylation, mutations).

Table 2: Sample of Growth and reduction

	ER+	ER-	Basal-like	Normal-like	HER2 enr	Luminal A	Luminal B
NA	219	81	21	31	38	119	28
Growth	29	6	1	4	2	14	5
Reduction	31	39	29	3	11	5	5

No association was discovered between mRNA and protein expression in group C, despite the influence of copy number levels on protein mRNA expression. Post-translational cleavage may be a possible explanation for this lack of connection in some proteins. Activation of caspase 7 and PARP1 are two examples of caspase-induced apoptosis, which can be activated by caspase 3 and calpain-1, respectively, in the context of necrosis and energy deprivation. mRNA expression levels were not controlled for in this investigation, hence only antibodies targeting the cleaved versions of these two proteins were used. This could have broken the association.

As seen in the remaining proteins (group D), the low correlation between all levels could be the result of a variety of reasons. Potential direct correlations between copy number and mRNA expression may be diminished by a significant fraction of non-aberrant malignancies. Indeed, when only tumor samples with copy number abnormalities were included in the analysis, a general rise in copy number to mRNA expression correlations was seen.

The mRNA-to-protein expression correlation would be impacted if both mRNA and protein degradation were to occur (in both group C and D). Group (C) and (D) proteins, on the other hand, have two or more subunits that are encoded by separate genes (two genes for GSK3 and seven for AMPK). Due to their linear link, proteins expressed by a single gene have a higher chance of reaching positive correlation. An untargeted antibody can also contribute to the noise. The EGFR antibody employed in this investigation also discovered high quantities of HER2 protein, as was the case with the EGFR. Due to inaccuracies in mRNA to protein expression correlations, EGFR data may be incorrect because of the nonspecific detection of Her2. Probe annotation is frequently updated, and erroneous matching of probes to genes and proteins may occur, despite the fact that the human genome sequence has been established. Splice variations can be detected by gene probes on the microarray which can tamper with comparisons (Rashid & Basusta, 2021). The increased correlations seen in the validation set may have been due to the improvement and optimization of antibodies and technologies. Correlations found between the two datasets were extremely significant, which supports the postulated regulatory linkages.

Variance

In order to achieve high correlation, there must be a systematic heterogeneity in the data. For group A proteins, the variance in copy number values was on average four times greater than for group B, C and D. Both mRNA and protein expression values for proteins in group B were the most variable. Since the expression range (signal distribution) of different proteins and genes might vary, the variance level is not directly comparable between proteins. Even though expression values (copy number, mRNA and protein expression) differed amongst subtypes, they were equivalent. For subtypes with similar or greater variance and low correlation, the high VSC indicates subtype-specific variations in regulatory mechanisms.

Tissue Heterogeneity

Within each tumor, there is a possibility of noise due to tissue heterogeneity. Because no microdissection was done, it's possible that different cell types contributed different amounts of DNA, mRNA, and protein to the sample. When comparing tumors as a whole, rather than individual tumor cells, researchers found relationships. Previous gene expression studies, however, found that tumor epithelial cells were the primary source of variance in mRNA expression associated with metastatic illness in DBCG 82 B and C cohorts when pathological information was taken into consideration (Aure, , et al., 2017). Despite the fact that the lysates were taken from the same tumor, they were not produced from the same cell type. Retrospectively, it would have been better if the tumor had been cut into three pieces and the DNA, mRNA, and protein isolated from each component. Intertumor protein levels were shown to be substantially more stable than intratumor protein levels in a recent technical review of RPPA data in comparison with gene expression. RPPA was able to accurately and reliably analyze protein expression despite the difficulties of intra-tumor heterogeneity (Bulut & Rashid, 2020).

miRNA regulation of cleaved caspase 7 expression

For cleaved caspase 7, high VSC was found for copy number to protein expression as well as mRNA expression to protein expression. Luminal B, HER2-enriched, and basal subtypes showed higher correlation values than luminal A, with the exception of the "normal-like" subtype. There are other possible explanations than differences in variance. There was a striking correlation between the mRNA levels of total caspase 7, suggesting that increased substrate levels may be a factor in the increased frequency of cleaved caspase 7. Although Calpain 1 (CAPN1) and caspase 3 (CASP3) cleave Caspase 7 protein, the basal-like subtype did not show greater expression levels of these two genes, which may explain the higher levels of cleaved Caspase 7 protein in this subtype. CASP7 is predicted to be a target of hsa-miR-29c, although the basal-like subtype of miRNA demonstrated its lowest expression. It is possible that the higher protein levels in basal-like cancers are due to a reduction in translational repression rather than a rise in CASP7 mRNA expression or an increase in cleavage activity by CAPN1 or CASP3. An analysis of miRNA expression in 101 early-stage breast carcinomas indicated that hsa-mir-29c and hsa-miR-29c* were among the top miRNAs that distinguished between basal-like and luminal A subtypes, further supporting their higher levels of expression in this subtype (Shan, , et al., 20197). In mesothelioma cell lines, higher levels of hsa-miR-29c expression have been linked to improved prognosis and lower proliferation, migration, invasion, and colony formation (Kumaran, , et al., 2017). MiRNAs have a tendency to be promiscuous in which mRNAs they may target, and their predicted "fine-tuning" regulatory actions can both repress translation and destabilize mRNA. For this reason (Rashid, et al., 2018), functional investigations are required to examine the precise effect of these miRNAs, as the biological signal-to-noise ratio is likely to be low.

Due to the fact that variance in the total group was broken into smaller, more uniform units, it was hypothesized that subdividing samples into molecular gene-expression subtypes would affect correlation. Because the variance in copy number and/or mRNA and protein expression values was reduced, subtyping penalized correlation. The statistical power was reduced due to the smaller sample sizes, and the increased number of tests led to an increase in false positives. Correlations between the ERBB2/HER2p1248 gene variants were found to have a substantial positive Z (high overall correlation across all samples and low mean subtype-specific correlation) (such as the HER2-enriched group). When it comes to CCNB1 copy number to mRNA correlation, the same process does not explain negative Z (low overall correlation and high mean subtype-specific correlation). The cyclin B1 copy number-to-mRNA expression association in the TCGA dataset had to be verified within each subtype due to the previously noted reduction in sample size and additional test issue. It was found that CCNB1 copy number levels did indeed influence the total expression level of CCNB1 in both the original and the validation sample sets. In contrast, it appeared that regulatory characteristics other than copy number were driving the majority of CCNB1 expression changes identified between subtypes. Including subtype-specific information in analyses of larger diverse datasets may obscure associations that otherwise would be obvious.

cyclin B1 was shown to be overexpressed in basal-like breast cancers at the gene and protein levels in a study of three cyclins (cyclin B1, D1, and E1) across molecular subtypes (Bareche, , et al., 2018). Cyclin B1 has been shown to play a significant role in the G2-M checkpoint and to be a biomarker for neoplasticity and aggressiveness in breast cancer, according to other research (Lin, , et al., 2018). The basal-like subtype of CCNB1 showed the most copy number reduction in our investigation. It is likely that the poor prognosis reported in the LOSS-group was caused by "basal-likeness," which was a confounder of CCNB1 copy number loss (Basudan, et al., 2019). CNB1 gain was found only in luminal-like and normal-like cancers, and it was found to have a somewhat worse prognosis in ER-positive tumors compared to the non-aberration group. This group's decreased survival may be attributable to the higher levels of CCNB1 expression that have developed as a result.

IV. CONCLUSION

Numerous molecular-level correlation studies have shown that the levels of DNA copy number, mRNA production, and protein production are all linked. We were able to classify genes and proteins into several regulatory pathways based on association patterns throughout such comparisons. Gene expression sub-type stratification was critical to uncovering previously unknown relationships at these levels. Correlation between subtypes was used to uncover putative subtype-specific regulation differences. Cleaved caspase 7 protein was shown to be related with reduced hsa-miR-29c expression. Basal-like tumors had a low overall survival rate, although the loss of CCNB1 copy number was not a probable cause. Gene copy number gain of CCNB1 was associated with an elevated cyclin B1 protein (and gene) expression and a worse prognosis in patients with luminary cancer.

REFERENCES

[1] Basudan, A., Priedigkeit, N., Hartmaier, R. J., Sokol, E. S., Bahreini, A., Watters, R. J., ... & Oesterreich, S. (2019). Frequent ESR1 and CDK pathway copy-number alterations in metastatic breast cancer. Molecular Cancer Research, 17(2), 457-468.

[2] Lin, C. Y., Beattie, A., Baradaran, B., Dray, E., & Duijf, P. H. (2018). Contradictory mRNA and protein misexpression of EEF1A1 in ductal breast carcinoma due to cell cycle regulation and cellular stress. Scientific reports, 8(1), 1-12.

[3] Bareche, Y., Venet, D., Ignatiadis, M., Aftimos, P., Piccart, M., Rothe, F., & Sotiriou, C. (2018). Unravelling triple-negative breast cancer molecular heterogeneity using an integrative multiomic analysis. Annals of oncology, 29(4), 895-902.

[4] Rashid, R. F., Çalta, M., & Başusta, A. (2018). Length-Weight Relationship of Common Carp (Cyprinus carpio L., 1758) from Taqtaq Region of Little Zab River, Northern Iraq. Turkish Journal of Science and Technology, 13(2), 69-72.

[5] Kumaran, M., Cass, C. E., Graham, K., Mackey, J. R., Hubaux, R., Lam, W., ... & Damaraju, S. (2017). Germline copy number variations are associated with breast cancer risk and prognosis. Scientific reports, 7(1), 1-15.

[6] Shan, W., Jiang, Y., Yu, H., Huang, Q., Liu, L., Guo, X., ... & Yang, Z. (2017). HDAC2 overexpression correlates with aggressive clinicopathological features and DNA-damage response pathway of breast cancer. American journal of cancer research, 7(5), 1213.

[7] Bulut, H., & Rashid, R. F. The Zooplankton Of Some Streams Flow Into The Zab River,(Northern Iraq). Ecological Life Sciences, 15(3), 94-98.

[8] Aure, M. R., Vitelli, V., Jernström, S., Kumar, S., Krohn, M., Due, E. U., ... & Sahlberg, K. K. (2017). Integrative clustering reveals a novel split in the luminal A subtype of breast cancer with impact on outcome. Breast Cancer Research, 19(1), 1-18.

[9] Craze, M. L., Cheung, H., Jewa, N., Coimbra, N. D., Soria, D., El-Ansari, R., ... & Green, A. R. (2018). MYC regulation of glutamine–proline regulatory axis is key in luminal B breast cancer. British journal of cancer, 118(2), 258-265.

[10] Rashid, R. F., & Basusta, N. (2021). Evaluation and comparison of different calcified structures for the ageing of cyprinid fish leuciscus vorax (heckel, 1843) from karakaya dam lake, turkey. Fresenius environmental bulletin, 30(1), 550-559.

[11] Schulz, D., Zanotelli, V. R. T., Fischer, J. R., Schapiro, D., Engler, S., Lun, X. K., ... & Bodenmiller, B. (2018). Simultaneous multiplexed imaging of mRNA and proteins with subcellular resolution in breast cancer tissue samples by mass cytometry. Cell systems, 6(1), 25-36.

[12] Zhao, Z. M., Yost, S. E., Hutchinson, K. E., Li, S. M., Yuan, Y. C., Noorbakhsh, J., ... & Yuan, Y. (2019). CCNE1 amplification is associated with poor prognosis in patients with triple negative breast cancer. BMC cancer, 19(1), 1-11.

[13] Rashid, R. (2017). Karakaya Baraj Gölünde (Malatya-Türkiye) yaşayan aspius vorax'da yaş tespiti için en güvenilir kemiksi yapının belirlenmesi/Determination of most reliable bony structure for ageing of aspius vorax inhabiting Karakaya Dam Lake (Malatya-Turkey).

[14] Sun, C. C., Li, S. J., Hu, W., Zhang, J., Zhou, Q., Liu, C., ... & Li, D. J. (2019). Comprehensive analysis of the expression and prognosis for E2Fs in human breast cancer. Molecular Therapy, 27(6), 1153-1165.

[15] Tanioka, M., Fan, C., Parker, J. S., Hoadley, K. A., Hu, Z., Li, Y., ... & Perou, C. M. (2018). Integrated analysis of RNA and DNA from the phase III trial CALGB 40601 identifies predictors of response to trastuzumab-based neoadjuvant chemotherapy in HER2-positive breast cancer. Clinical Cancer Research, 24(21), 5292-5304.

[16] Bulut, H., Rashid, R. F., & Saler, S. Erbil (Irak) İlinde Bulunan Bazi Göletlerin Zooplanktonu Öz.

[17] Jung, H., Kim, H. S., Kim, J. Y., Sun, J. M., Ahn, J. S., Ahn, M. J., ... & Choi, J. K. (2019). DNA methylation loss promotes immune evasion of tumours with high mutation and copy number load. Nature communications, 10(1), 1-12.

[18] Guha, M., Srinivasan, S., Raman, P., Jiang, Y., Kaufman, B. A., Taylor, D., ... & Avadhani, N. G. (2018). Aggressive triple negative breast cancers have unique molecular signature on the basis of mitochondrial genetic and functional defects. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 1864(4), 1060-1071.

[19] Pala, G., Caglar, M., Faruq, R., & Selamoglu, Z. (2021). Chlorophyta algae of Keban Dam Lake Gülüşkür region with aquaculture criteria in Elazıg, Turkey. Iranian Journal of Aquatic Animal Health, 7(1), 32-46.

[20] Rashid, Rf, Çoban, Mz, & Saler, S. Evaluation Of Water Quality Of Keban Dam Lake (Elaziğ-Turkey).

[21] Hasna, J., Hague, F., Rodat-Despoix, L., Geerts, D., Leroy, C., Tulasne, D., ... & Kischel, P. (2018). Orai3 calcium channel and resistance to chemotherapy in breast cancer cells: the p53 connection. Cell Death & Differentiation, 25(4), 693-707.

[22] Rashid, R. F., & Saler, S. Effects Of Global Warming On Aquatic Life.

[23] Naidoo, K., Wai, P. T., Maguire, S. L., Daley, F., Haider, S., Kriplani, D., ... & Natrajan, R. (2018). Evaluation of CDK12 protein expression as a potential novel biomarker for DNA damage response–targeted therapies in breast cancer. Molecular cancer therapeutics, 17(1), 306-315.

[24] Shekha, M. S., Hassan, A. O., & Othman, S. A. (2013). Effects of Quran listening and music on electroencephalogram brain waves. Egypt. J. Exp. Biol, 9(1), 1-7.

[25] Buccitelli, C., & Selbach, M. (2020). mRNAs, proteins and the emerging principles of gene expression control. Nature Reviews Genetics, 21(10), 630-644.

[26] Xiao, B., Chen, L., Ke, Y., Hang, J., Cao, L., Zhang, R., ... & Li, L. (2018). Identification of methylation sites and signature genes with prognostic value for luminal breast cancer. BMC cancer, 18(1), 1-13.

[27] Johnstone, C. N., Pattison, A. D., Gorringe, K. L., Harrison, P. F., Powell, D. R., Lock, P., ... & Anderson, R. L. (2018). Functional and genomic characterisation of a xenograft model system for the study of metastasis in triple-negative breast cancer. Disease models & mechanisms, 11(5), dmm032250.