Preprints.org

Article

Integrated Bioinformatics Analysis for the Screening of Hub Genes in Breast Cancer Metastasis Recurrence at Liver

Tyler Kwok , Suneetha Yeguvapalli , Kumaraswamy Naidu Chitrala [*]

Article

Integrated Bioinformatics Analysis for the Screening of Hub Genes in Breast Cancer Metastasis Recurrence at Liver

Tyler Kwok, Suneetha Yeguvapalli and Kumaraswamy Naidu Chitrala *

Department of Engineering Technology, University of Houston, Sugar Land, TX 77479, USA

* Correspondence: kchitral@central.uh.edu; Tel.: +1 7137437965

Abstract: Breast cancer, when advancing to a metastatic stage, involves the liver, impacting over 50% of cases and significantly diminishing survival rates. Presently, a lack of tailored therapeutic protocols for Breast Cancer Liver Metastasis (BCLM) underscores the need for a deeper understanding of molecular patterns governing this complication. Therefore, by analyzing differentially expressed genes (DEGs) between primary breast tumors and BCLM lesions, it aimed to shed light on the diversities of this process. This research investigates Breast cancer liver metastasis relapse by employing a comprehensive approach that integrates data filtering, Gene Ontology and KEGG Pathway Analysis, Overall Survival analysis, identification of the Alteration in the DEGs, Visualization of Protein-Protein Interaction Network, Signor 2.0, identification of positively correlated genes, screening results of the function of hub genes, immune cell infiltration analysis, genetic alternation analysis, Copy number variant analysis, Gene to mRNA interaction, transcription factor analysis, autodocking and identification of Potential Treatment Target. This study's integrative approach unveils metabolic reprogramming, suggesting altered *PCK1* and *LPL* expression as key in breast cancer metastasis recurrence.

Keywords: breast cancer recurrence; Breast Cancer Liver Metastasis; Bioinformatic analysis
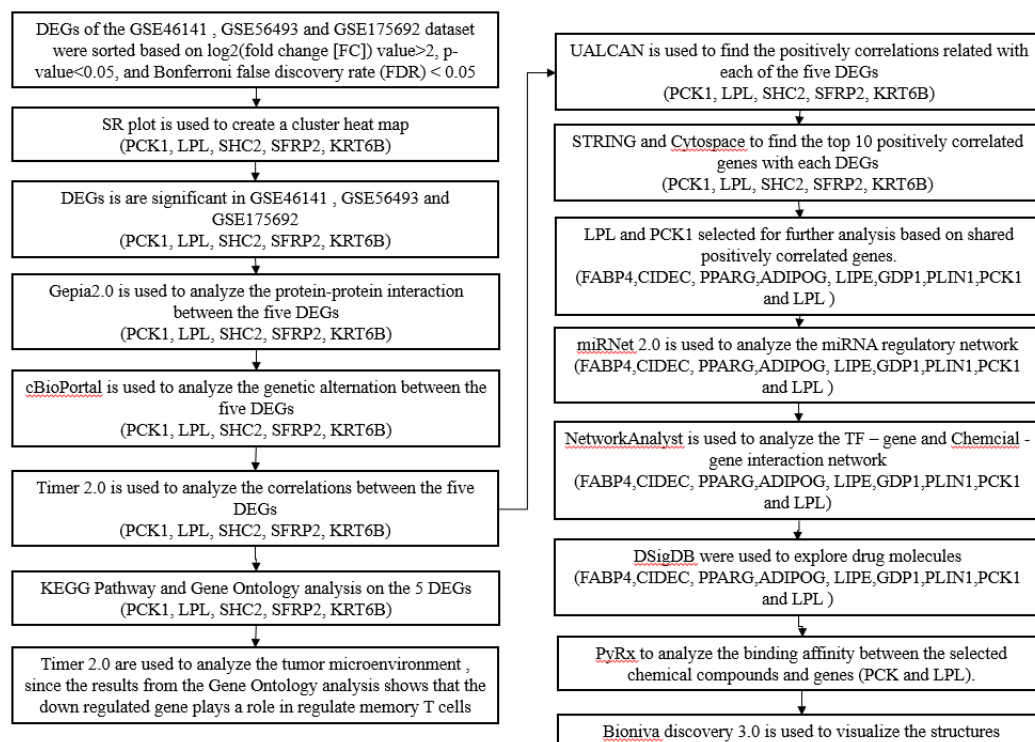
1. Introduction

Liver metastasis is a common and devastating complication for patients with breast cancer, occurring in over 50% of those with advanced disease. Though early-stage breast cancer has a 5-year survival rate approaching 100%, this prognosis drastically declines by 26% upon diagnosis of distant organ involvement [1]. Unfortunately, there are currently no standardized therapeutic protocols explicitly tailored for recurrent breast cancer liver metastasis (BCLM) due to an incomplete understanding of the molecular patterns and mechanisms governing this metastatic process. Elucidating the genetic factors that drive BCLM development and relapse following treatment is imperative to determine patient prognosis, identify the highest risk, and develop targeted therapies.

A critical first step is to define the landscape of differentially expressed genes (DEGs) between primary breast tumors and matched BCLM lesions. However, uncertainty remains regarding the key molecular biomarkers and functional pathways involved in BCLM pathogenesis. A thorough investigation of DEGs between liver metastases and the original cancer can thus provide clinically valuable information on the prognostic and therapeutic potential of these genetic alterations. In summary, comprehensive profiling of differential gene expression patterns between primary and metastatic liver tumors can facilitate prognosis prediction, inform clinical decision-making, and reveal novel targets and strategies for treating this common yet poorly understood complication of advanced breast cancer.
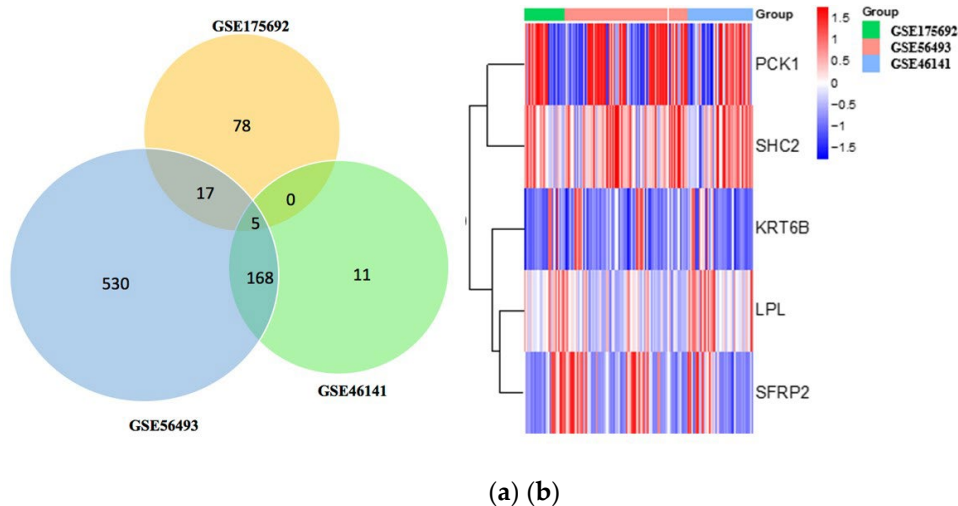
2. Results

2.1. Identification of Datasets and Analysis of Differentially Expressed Genes

`The overall summary of the data analysis conducted in the study is shown in Figure 1. Gene expression profiles from the Gene Expression Omnibus (GEO) datasets GSE46141, GSE56493, and GSE175692 from breast cancer patients (n=125) were selected for our study. Among these samples, pre-treatment ones (n=83) were selected for our study. Results from the differential analysis between breast cancer local metastasis (n=16) versus breast cancer liver metastasis (n=11) in the GSE46141 dataset showed that 184 were significantly expressed with 154 genes being up-regulated and 30 genes being down-regulated. DEG analysis of the breast cancer metastasis samples from the breast (n=19) versus the liver (n=27) in the GSE56493 dataset showed that there were 720 genes significantly expressed with 434 being upregulated and 286 being downregulated. Further, DEG analysis of relapsed metastatic breast cancer in the breast (n=16) versus liver (n=27) showed that there were 100 genes was significantly expressed with 59 genes being up-regulated and 41 genes being down-regulated. Results from Venn analysis demonstrated that there were 5 genes (PCK1, SHC2, LPL, SFRP2, KRT6B) overlapped between the three datasets (Figure 2a). Among the 5 genes, two genes (PCK1, SHC2) were upregulated and three (LPL, SFRP2, KRT6B) were downregulated. The expression levels of these DEGs were provided in the heatmap below (Figure 2b).
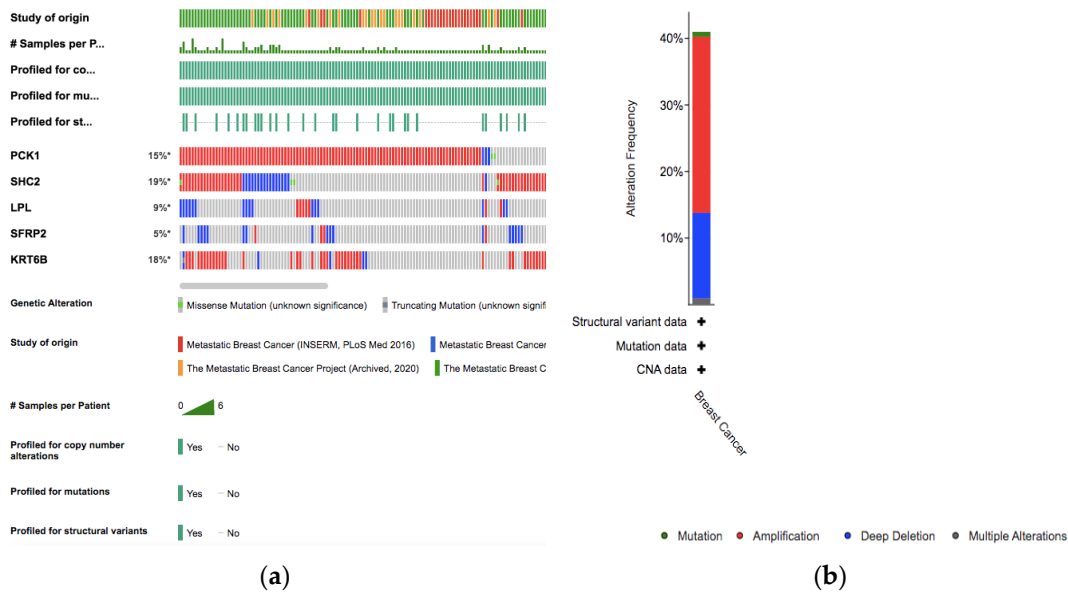

Figure 1. Flow chart of the bioinformatics approaches used in this study.
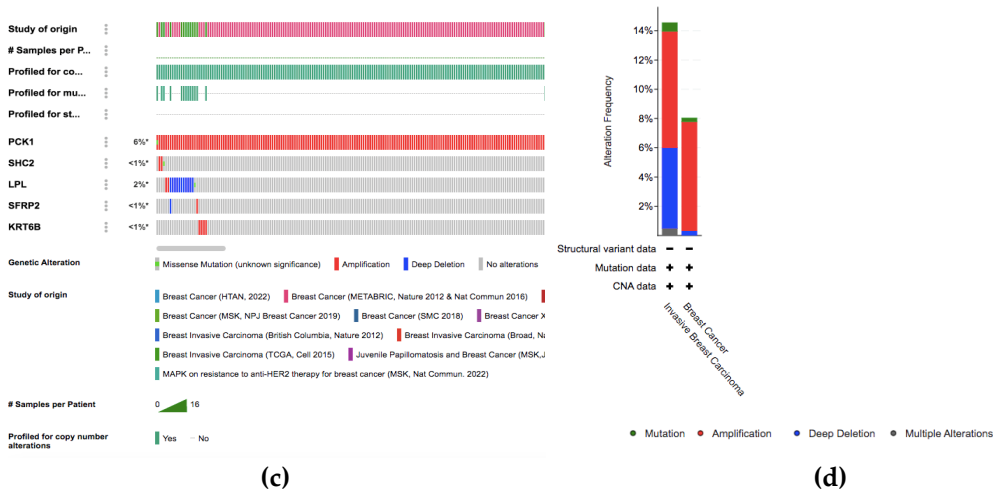
Figure 2. Analysis of differentially expressed genes **a)** Venn analysis of DEGs in the three datasets (GSE175692, GSE 56493, and GSE 46141) (**b**) Heatmap for the genes that are differentially expressed and common between the datasets GSE175692, GSE 56493, and GSE 46141.

2.2. Identification of Genetic Alterations in the DEGs

For the five genes (LPL, PCK1, KRT6B, SFRP2, SHC2) that are differentially expressed between the datasets, we analyzed the alterations in Metastatic Breast cancer and Invasive breast cancer genetic profiles (Figure 3). Results from the metastatic breast samples showed that the SHC2 gene has the highest alterations in the genetic biomarker at 19%, KRT6B has alterations at 18%, and PCK1 has alterations at 15%. LPL(9%) and SFRP2(5%) both have lower genetic alterations compared to the other DEGS. In all the genes, many alterations are mostly composed of amplification and deep deletion. Conversely, there are no significant results shown in Breast Cancer Invasive dataset. PCK1 has a alterations at 6% and all the other genes have a genetic alterations lower than 2%.

(c)   **(d)**
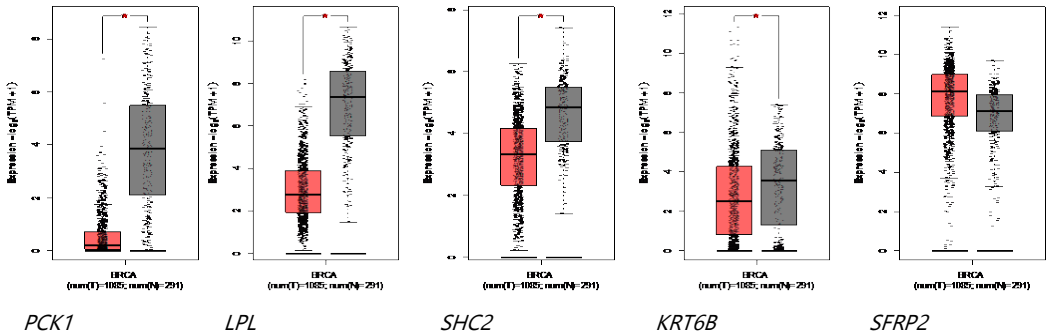
Figure 3. Analysis of genetic alterations related to the 5 DEGs across the genomic data sets (A) from Breast Cancer Metastasis dataset [2197 samples/1813 patients] (B) is an overall summary of alternations of the 5 DEGs in genomic data set of Breast Cancer Metastasis (C) from Breast Cancer Invasive data [4184 samples/4046 patients] (D) is an overall summary of alternations of the 5 DEGs in genomic data set of Breast Cancer Invasive.

2.3. Prognostic Information Hub Gene Expression

The Overall survival and Copy number variants of the DEGs were analyzed using the Breast Cancer integrative platform. The Kaplan-Meier plotter (Supplementary Figure 1) was used to identify the survival of the 5 DEGs. Results showed that the expression level SFRP2[HR = 0.544, P = 3.08e-5], SHC2 [HR = 6.54e-6, P = 0.733]and PCK1[HR = 0.541, P = 0.00175], was positively correlated with the overall survival of patients. The expression level of LPL[HR = 0.843, P = 0.0126] and KRT6B [HR = 0.736, P = 0.00787] doesn't show a significant correlation to the overall survival of patients. Additionally, we analyzed the expression of the DEGs for copy number variants (CNV) (Supplementary Figure 2) gain (amplifications) and loss (deletions). Results showed that the expression of PCK1 is high when CNV is gained. Conversely, there is a high expression of LPL, SHC2, and SFRP2 when CNVs are lost. There is no significant change in expression in CNV for KRT6B. Then, compared the gene expression in CNV gain and loss to the survival analysis, but there isn't any significant impact for all of the genes.

The GEPIA2 database was used to verify the expression levels of the DEGs in tumor and normal tissues. As shown in Figure 4, SHC2, LPL, PCK1, and KRT6B are all under-expressed, and only SFRP2 was highly expressed in the breast tumor. Additionally, the expression levels of the DEGs in different stage BC patients are shown in the Supplementary Figure 3. As detail listed in the Supplementary Table 1, PCK1, LPL, and SFRP2 play a role in breast cancer's initial and developmental stages, and KRT6B is exclusively related to the developmental stages.

PCK1   LPL   SHC2   KRT6B   SFRP2

Figure 4. Expression of Genes in Breast Invasive Carcinoma; Red bars represent expression in Cancer tumors and the grey bars represent expression in normal patients. * represents statistical significance at a p-value < 0.05.

2.4. Gene Network Analysis

To analyze the Protein-Protein interaction of the DEGs we used Gene mania (Figure 5). *LPL, PCK1*, and *SHC2* are co-expressed, and there is a genetic interaction between *LPL* and *PCK1*. We use Signor 2.0 to visualize the signaling pathways between the co-expressed genes with a confidence score greater than 0.5 (Supplementary Figure 4). We validate the gene relationship between *LPL* and *PCK1* with Timer 2.0 (Supplementary Figure 5A). Both genes have a significant positive correlation in all breast Subtypes. The result also shows that downregulated genes *SFRP2* and *LPL* positively correlate in all breast cancer subtype groups (Supplementary Figure 5B). Analysis conducted through Gepia 2.0 futher demonstrate the expression of *SFRP2* and *LPL* in different breast cancer subtype groups (Supplementary Figure 6). Results shows that *LPL* is still under-expressed in tumor cells and *SFRP2* is overexpressed in all HER2, Luminal A, and Luminal B subtype groups and under-expressed in the Basal_like subtype group.

Figure 5. Protein-protein interaction analysis of DEGs. Red line represents physical interaction, purple line represent co-expression, orange line represent predicted, blue line represent co-localization, Green represent genetic interaction and light blue represent pathways.

2.5. Identify Positive Correlated Genes

To gain insights into potential interacting partners and biological pathways associated with the DEGs, a series of integrated bioinformatics analyses were performed. First, the UALCAN online tool was used to identify genes exhibiting a positive Pearson correlation (coefficient >0.4) with the expression levels of the 5 DEGs across breast cancer TCGA datasets. Next, the resultant gene list was uploaded to the STRING database to construct a protein-protein interaction (PPI) network (Supplementary Figure 7) incorporating the DEGs and their correlated genes. Cytoscape software with the CytoHubba application was then used to analyze the topology of the STRING generated PPI network and selected the top 10 hub genes for *SHC2 (PHPT1, MRPL41, ATP5F1D, RHOT2, NME3, TUBGCP6, SNRNP70, WDR90, THEM259, E4F1), PCK1 (PPARG, ACACB, PLIN1, LIPE, ADIPOQ,*
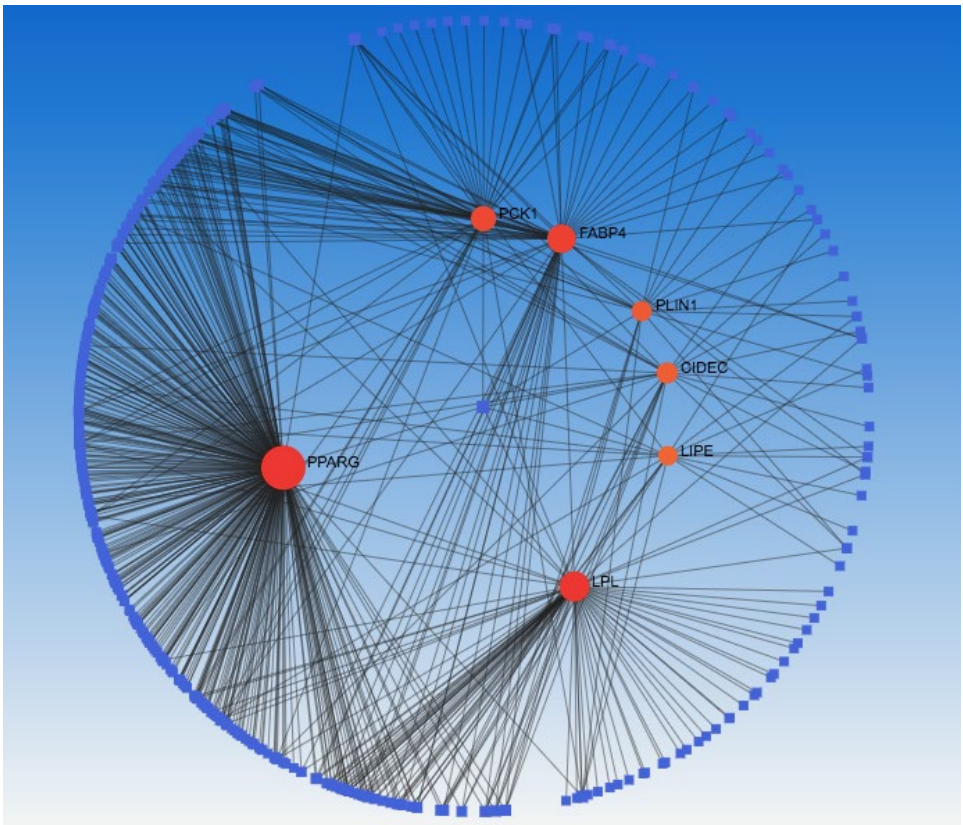
GPD1, CIDEC, PCK1, LPL, FABP4), LPL (PPARG, ADIPOQ, CAV1, LIPE, LPL, PLIN1, CIDEC, FABP4, CD36, GDP1), KRT6B (TRIM29, KRT14, KRT15, PKP1, KRT6B, DSG3, KRT16, KRT17, SERPINB5, DSC3), SFRP2 (COL1A1, MMP2, DCN, LUM, TGFB1, BGN, COL1A2, POSTN, FN1, COL3A1) ranked by their degree of interaction (Supplementary Figure 8A). Gene Ontology (GO) term (Supplementary Figure 8B) and KEGG pathway (Supplementary Figure 8C) enrichment analyses of these 10 hub genes were conducted using the DAVID database. This revealed that two selected genes, phosphoenolpyruvate carboxykinase 1 (*PCK1*) and lipoprotein lipase (*LPL*), were commonly enriched in pathways involved in fatty acid and energy metabolism regulation, including PPAR signaling, AMPK signaling, regulation of lipolysis in adipocytes, and thermogenesis. Integrating interactome and pathway databases in this manner provided novel biological insights into molecular mechanisms potentially implicated in breast-to-liver metastasis by DEGs and their correlated network hubs (*FABP4, CIDEC, PPARG, ADIPOG, LIPE, GDP1, PLIN1, PCK1 and LPL*).

2.6. Gene to miRNA and Transcription Factor Interaction

A network were constructed to analyze the relationship between the target miRNA and Transcription factors with the network hub genes (*FABP4, CIDEC, PPARG, ADIPOG, LIPE, GDP1, PLIN1, PCK1* and *LPL*). miRNet 2.0 was applied to screen the targeted miRNAs of the hub genes. As illustrated in Supplementary Figure 9, the interaction network consists of 70 nodes and 88 edges. The interactive hub genes that most miRNAs target are LPL (degree score = 30) and PPARG (degree score = 29) and the highest interactive miRNAs is hsa-mir-27a-3p (degree score = 5). From the results we also identified four miRNAs target both the PCK1 and LPL. miRNA (hsa-mir-27a-3p) targeted *LPL, PCK1, CIDEC, PLIN1* and *PPARG*; miRNA (hsa-mir-124-3p) targeted *LPL, PCK1* and *CIDEC*; and both miRNA (hsa-mir-10b-5p and hsa-mir-200b-3p) targeted *LPL* and *PCK1*. Transcription factor (TF) enrichment analysis was performed to identify overrepresented TF binding sites in promoters of the hub genes using network analyst (Supplementary Figure 10). Results showed that the transcription-regulated network of hub genes included 30 nodes and 33 edges. *NFKB1* and *RELA* are the highest interactive transcription factor with a degree of 3. Among the transcription factors, SP1 is the only regulator that was common in both *LPL* and *PCK1* and it mainly plays a role in transcriptional mis-regulation in cancer.

2.7. Identification of Potential Treatment Targets

A network analysis is performed to identify the chemical to hub genes interaction (Figure 6). Results showed that there are 447 nodes and 634 edges with bis(4-hydroxyphenyl)sulfone, Dexamethasone and rosiglitazone has the highest degree (7) of interaction within the hub genes. To drug targets for the hub genes, we submitted them to DSigDB database (Table 1) and ranked them based on their Adjusted P-value. Our results showed that Triflumizole and Rosiglitazone CTD are the most significant drugs of the hub gene in the DSigDB database, followed by IBMX BOSS, formic acid BOSS, IBMX CTD 00007018, Bisphenol A diglycidyl ether CTD 00000976, oleic acid BOSS,glycerol BOSS, D-glucose BOSS and insulin BOSS.

Figure 6. Identification of interaction between hub genes and chemicals. Network showing the hub genes are shown in red color nodes and whereas chemicals shown in blue color nodes. Line refers to the interaction between the hub genes and chemical.

Table 1. Potential drug treatment targets with clinical application identified for hub genes.

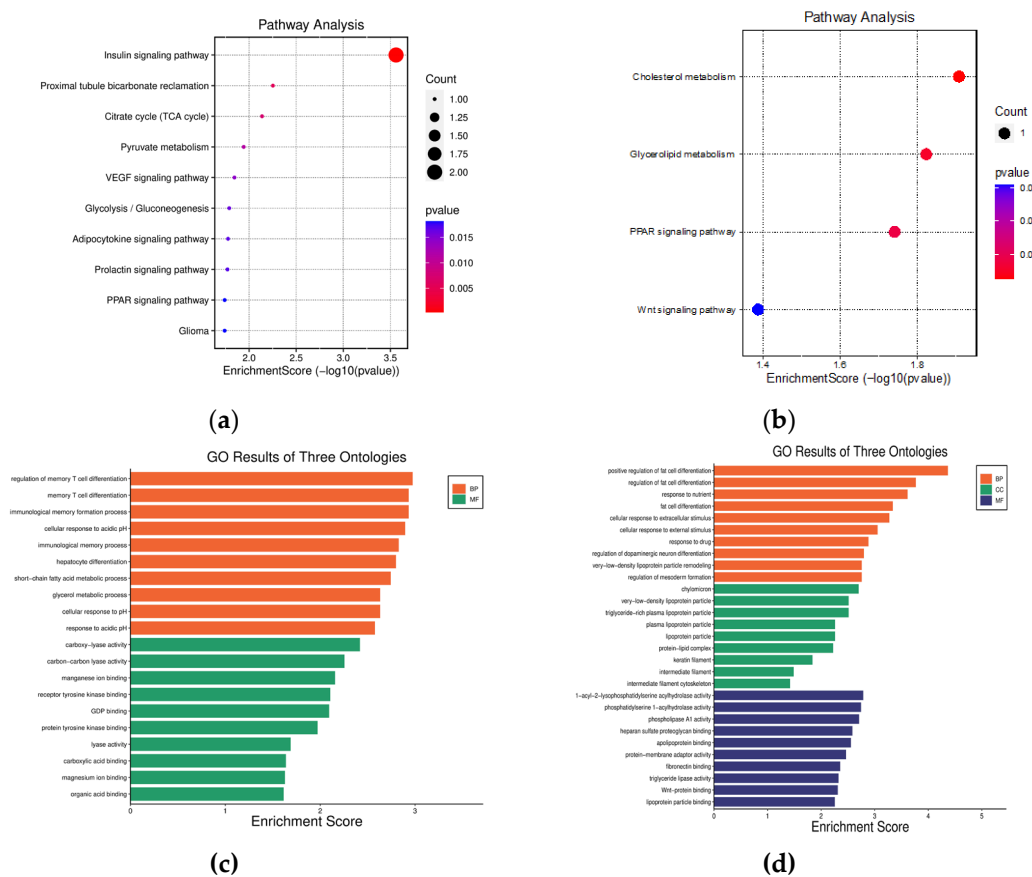| Name	P-value	Adjusted p value	Combined score
Triflumizole CTD 00002280	3.728e-15	3.158e-12	118549.00
Rosiflitazone CTD 00003139	2.628e-13	1.110e-10	11263.85
IBMX BOSS	2.049e-12	5.771e-10	20983.21
Formic acid BOSS	4.897e-11	1.034e-8	5525.67
IBMX CTD 00007018	1.826e-10	3.086e-8	5723.06
BISPHENOL A DIGLYCIDYL ETHER CTD 00000976	8.713e-10	1.227e-7	11104.17
Oleic acid BOSS	3.661e-9	4.419e-7	3127.80
Glycerol BOSS	7.426e-9	7.492e-7	2604.39
D-glucose BOSS	8.503e-9	7.492e-7	2514.27
Insulin BOSS	9.207e-9	7.492e-7	2462.76

2.8. Molecular Docking for Protein-Chemical Interaction

The 4 common chemicals (bis(4-hydroxyphenyl)sulfone, Dexamethasone, Triflumizole and Rosiglitazone) that showed highest interaction between the hub genes were used to analyze their respective interactions with the genes LPL and PCK1. Our results (Supplementary Table 2) showed that Rosiglitazone has the highest binding affinity(-8.4) with 1NHX formed three hydrogen bond at THR A:339, ASN A:292 and VAL A:335, three pi-alkyl interaction are seen at LYS A:290,VAL A:335 and ARG A:87, and one pi-cation and one Carbon Hydrogen bond is formed at ARG A:405 and ASP A:311 (Supplementary Figure 11b). Dexamethasone (-8.1) showed second highest binding affinity towards 1NHX. Dexamethasone formed 2 hydrgon bonds at ARG A:405, and one hydrogen bond at

LYS A:290 and CYS A:288. There is also a Halogen (Fluorine) at ARG A:405(Supplementary Figure 11c). Bis(4-hydroxyphenyl)sulfone(-7.5) formed 5 hydrogens bond at HIS A:264, ASP A:311, SER A:286, AIA A:287 and CYS A:288, one pi-Alkyl and Pi sigma interaction was formed at LYS A:290 and VAL A:335 respectively (Supplementary Figure 11d). Triflumizole (-7.0) formed two hydrogen bonds with PHE A:530 and ASN A:533. Pi-pi T-shaped formed bonding with PHE A:525 and pi-pi stacked formed bonding with PHE A:530, one flourine interaction and carbon hydrogen bond was formed with CYS A:288 and TRP A:516 resepctively. Four pi-Alkyl interaction are formed at TRP A:516,PHE A:517,PHE A:525 and PHE A:530 (Supplementary Figure 11a). Rosiglitazone(-7.7) also has the highest binding affinity when it interact with 6e7k, following with Triflumizole (-7.5), Dexamethasone (-7.3), bis(4-hydroxyphenyl)sulfone (-7.3). As seen in supplementary Figure 12b, Rosiglitazone has three pi-alkyl interaction at ILE A:221, LYS A:265 and VAL A:84, one pi-pi stacked, pi-pi t-shaped, pi-cation and cabron hydrogen bond are formed at TYR A:121,TRP A:82, LYS A:265 and HIS A:268 resecpectively. Triflumizole (Supplementary Figure 12a) has two halogen(Flourine) at ILE A:221, HIS A:268 and two hydrogen bond at SER A:159, HIS A:268. Six pi-alkyl are formed with PHE A:212, HIS A:268,LYS A:265 and ILE A:221 and twice at TYR121, five alkyl are seen at, VAL A:264, PRO A:187 and twice at ILE A:221. Three Carbon Hydrogen bond are formed with HIS A:268, SER A:159, PRO A:187. Dexamethasone and bis(4-hydroxyphenyl)sulfone both have the same binding affinity. Dexamethasone formed two hydrogen bonds at at THR A:379 and GLU A:298. One Halogen (Florine) and unfavorable acceptor- acceptor interaction at GLU A:298 and SER A:390 respectively(Supplementary Figure 12c). Bis(4-hydroxyphenyl)sulfone formed two hydrgon bond with HIS A:268 and TYR A:158, two Pi-Pi t-shaped are form with TRP A:82. One Pi-cation and Pi-Sulfur was formed at LYS A:265 and HIS A:268. Two Pi-Alkyl interaction was formed at LYS A:265 and ILE A:221(Supplementary Figure 12d). All the chemical compounds were tested for their drug-likeliness. Results showed that these compounds have druglikeness (supplementary Figure 13).

2.9. Gene Ontology (GO) and Functional Enrichment Analysis

The Database for Annotation, Visualization, and Integrated Discovery (DAVID; version 6.8) was utilized to perform the functional enrichment analysis of the five most prevalent DEGs. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis identified the molecular pathways enriched among the up-and down-regulated genes. The 2 upregulated DEGs were significantly enriched in the insulin signaling pathway(Figure 7A). Meanwhile, the 3 downregulated DEGs were expressed in several metabolic pathways(Figure 7B), including Cholesterol metabolism, Glycerolipid metabolism, and Peroxisome proliferator-activated receptor (PPAR) signaling pathway. Gene Ontology (GO) term enrichment was also conducted across biological processes, cellular components, and molecular functions. The upregulated DEGs (Figure 7C)were primarily associated with regulating the memory T-cell differentiation for biological processes. Considering that the up-regulated genes (SHC2 and PCK1) have a role in regulation of memory T cell differentiation and immunological memory formation process, we examined the correlations the up regulated genes and infiltrated immune cells. Base on supplementary graph 12 we can see that both SHC2 and PCK1 are positively correlated with the infiltrating levels of NK cell, Mast cell activated and the macrophages M2. In contrast, downregulated DEGs (Figure 7D) were mainly involved in the positive regulation of adipocyte differentiation from precursor cells, response to nutrients, and cellular response to extracellular stimuli. These results from DAVID highlight the dysregulation of insulin signaling and metabolic pathways that may contribute to the pathobiology of breast cancer liver metastasis according to expression patterns of the consistent DEGs between datasets.

Figure 7. Function enrichment analysis for DEGs. KEGG pathway analysis with up-regulated (A) and down-regulated (B) DEGs. GO analysis with up-regulated (C) and down-regulated (D) DEGs.

3. Discussion

Recent studies and research indicate that breast cancer is currently the most frequent type of cancer among females, exceeding "lung cancer as the leading cause of cancer incidence" in the US and globally [1]. Roughly 20% of individuals diagnosed with breast cancer will encounter a recurrence and 50–70% of metastatic breast cancer cases involve the liver [2]. Moreover studies have shown that there is a poor prognosis for breast cancer metastasis to the liver, with the median survival rate being only 2–3 years [3]. Although breast cancer liver metastasis (BCLM) is relatively common, there is a scarcity of specialized therapeutic options available for patients. This signifies that there is a dire need for the development of interventions in the area, which could only be made possible by rigorous research and an in-depth understanding of molecular processes underlying breast cancer to allow progress in early diagnosis and treatment of the condition [4]. Therefore our bioinformatic research will identify potential biomarkers and assist in predicting the accurate behavior of the illness, and aid in the creation of targeted therapeutic approaches. Researching biomarkers can provide insight into the molecular processes that underlie the spread of cancer, its metastasis, and its resistance to treatment. This information can further our understanding of cancer biology and aid in the creation of novel targeted treatments.

In this study, we used proven online bioinformatics tools to investigate possible biomarkers for diagnosis of breast cancer and therapeutic drugs. We identified 5 DEGs common to all three GEO datasets, which included 2 upregulated genes and 3 downregulated genes. The upregulated genes were mainly involved in insulin signaling pathway, which is an important factor for breast cancer prognosis [5]. The downregulated genes didn't have significant results in pathways enrichment, but the gene ontology results show that they play a role in regulation of fat cell differentiation. Genetic alteration analysis was performed on the DEGs in metastatic breast samples. We observed major genetic alterations in *PCK1,SHC2* and *KRT6B*, emphasizing the role of these genes in tumor behavior

modification as well as response to treatment. These changes, such as amplification and deep deletion remain significant for understanding the complex mechanisms of breast cancer evolution. In addition, the CNV analysis shows adaptive cellular response to genetic variations. With the assistance of BCIP platform, we found that the expression of *PCK1* is higher when CNV is gained. This may imply the increase of expression changes the cancer cell metabolism. When tumors are coming from different organs, the combination of genetic alteration and environmental stress would determine the expression level of *PCK1*. For instance, a high expression of gene *PCK1* in CNV gain influences the inducing of retrograde carbon flow from gluconeogenesis [6]. As a result, there would be a decrease in glutathione levels while ROS production would be enhanced, thereby supporting hypoxic breast cancer growth. Since *PCK1* performs an anti-oncogenic role in gluconeogenic organisms, its gains in CNV would thus affect breast cancer metastasis in the liver due to its tumor-promoting roles. Specifically, *PCK1* enhances liver metastatic growth by facilitating pyrimidine nucleotide biosynthesis in hypoxia, a significant feature in the liver microenvironment. Reversely, the *SHC2* expression is higher after CNV loss suggests but previous studies have shown that high expression of *SHC2* in CNV loss does not affect breast cancer metastasis in the liver [7]. Ideally, high expression of *SHC2* does not lead to significant variations associated with human neurological conditions, therefore the change in gene *SHC2* copy number is not vivid in breast cancer or liver.

To gain further insight on the functions and potential role that the 5 DEGS have in Breast cancer Local metastasis and in Metastasis in liver, we collected genes that were positively related with them (with a score >0.4). We created a PPI among all the positively correlated genes for each DEGs and identify the hub genes that can be used for Gene ontology and KEGG pathway enrichment analysis. When, we analyze each DEGs and their positively correlated hub genes, we found that the Pathway analysis and Gene Ontology results are closely related between *PCK1* and *LPL*. Therefore, we took the Top degree shared hub genes between *PCK1* and *LPL* to explore and create network analysis. It is controversial whether *PCK1* plays an oncogenic or tumor suppressor function in various human cancers. *PCK1* has antitumorigenic effects in gluconeogenic organ cancers (liver and kidney) but tumor-promoting effects in non-gluconeogenic organ malignancies. Prior studies have shed light on *PCK1*'s hijacking function and mechanisms in cancer of the colon, hepatocellular carcinoma, breast, kidney, etc. Higher *PCK1* activity allows metastatic breast cancer cells to engage in gluconeogenesis, thereby generating glucose and glyceroneogenesis [8]. This allows for the creation of biosynthetic operations. This metabolic adaptation enables the cells to survive in a liver microenvironment. *PCK1* expression allows breast cancer cells that have metastasized to the liver to disadvantage, driving key metabolic fluxes. *LPL* plays a major role in influencing Breast cancer liver metastasis because it affects the rates of lipid metabolism. Breast cancer is influenced by metabolic abnormalities that affect its onset and progression, especially increased lipid synthesis and uptake. However, depending on the patient's hormone status and the stage of the disease, lipid metabolism has different effects on breast cancer. The maintenance of the proliferation and survival of breast cancer cells depends on the dysregulation of fatty acid metabolism [9]. Liver cancer cells are more likely to take up and use fatty acids when *LPL* expression is higher. As a result, there is an increased potential for breast cancer cells to proliferate and spread into liver tumors.

In our miRNA and hub gene analysis, we identified four miRNA of interest, which include hsa-miR-27a-3p, hsa-miR-124-3p, hsa-miR-10b-5p and hsa-miR-200b-3p. These targeted miRNA are important regulators involved in the development and metastasis of breast cancer that are mainly associated with the change in expression of *PCK1* and *LPL*. Additionally, has-miR-27a-3p has been found to be a two-sided regulator of transformation of metabolism and metastasis [10]. Evidence has shown that it can reduce the expression of *PCK1*, which is an essential enzyme in gluconeogenesis and shift the metabolism toward favoring cell proliferation and metastasis. The silencing of the *PCK1* gene by hsa-miR-27a-3p may trigger a metabolic reprogramming of cancer cells which would make them to take over and become more resistant and invasive when under nutrient starvation, which is a typical situation in the tumor microenvironment.

On the other hand, it was found that hsa-miR-124-3p functions as a tumor suppressor in various cancers such as breast cancer. It has been demonstrated that *LPL* downregulation and targeting by

this drug is critical for cancer cells liposuction which is crucial for providing energy and molecules required for their fast growth and reproduction. Hsa-miR-124-3p can block the *LPL* function and therefore inhibit the lipid metabolism and the cell energy homeostasis of the breast cancer cells which would be a strategy to control its metastasis [11]. Hsa-miR-10b-5p, which is famous for pro-metastatic effects, was suggested as a *PCK1* regulation directly. Over-expression of the precursor of hsa-miR-10b-5p is directed towards the downregulation of *PCK1* activating glycolysis over gluconeogenesis, enhancing the Warburg effect, a metabolic hallmark of cancer cells and facilitating metastasis of breast cancer cells.

Moreover, hsa-miR-200b-3p, a member of the miR-200 family, plays a crucial role in controlling Epithelial-Mesenchymal Transition (EMT), a key process in cancer metastasis progression [12]. While there is not a direct link between *PCK1* and *LPL* expressions, they could contribute to EMT, thereby promoting metabolic changes and lipid metabolism. These alterations may enhance the metastatic capability of breast cancer cells.

A TFs network of hub genes was constructed to explore the molecular networks associated with Breast cancer. From the network, Sp1 is the only transcription factor that is shared between *PCK1* and *LPL*, it is mainly involved in Transcriptional mis regulation in cancer. Studies have shown that Sp1 can directly bind to the promoter region of the *PCK1* [13] and controls the proliferation of breast cancer cells by interacting with insulin-like growth factors I receptor [14], enhancing their transcriptional activity. Therefore, targeting Sp1-mediated transcriptional regulation may be a potential therapeutic strategy, but more investigations are needed to understand TFs and hub gene interactions in regulating Breast Cancer liver metastasis.

We did a chemical-to-gene interaction of the nine hub genes that were identified in the studies. Rosiglitazone, Dexamethasone, and bis(4-hydroxyphenyl)sulfone were shown to interact closely with the hub genes. Rosiglitazone, a pharmacological compound, binds to PPARγ (peroxisome proliferator-activated receptor gamma), a nuclear receptor protein controlling gene expression [15]. Activation of PPARγ by Rosiglitazone leads to the overexpression of several genes crucial in lipid and glucose metabolism. For instance, it triggers increased activity of *LPL*, encoding enzymes to break down triglycerides from lipoproteins, thereby enhancing their breakdown in the blood. Additionally, Rosiglitazone induces the expression of *PCK1*, encoding enzyme in the pathway producing glucose from non-carbohydrates, ultimately enhancing hepatic glucose output. Further analysis in drug databases revealed Rosiglitazone CTD 00003139 as a potential treatment, validating the importance of Rosiglitazone in treatment of recurrent breast cancer metastasis in the liver. Further research are needed to be done for Dexamethasone and bis(4-hydroxyphenyl)sulfone in Breast cancer liver metastasis recurrence. The precise effect of bis(4-hydroxyphenyl)sulfone on breast cancer is remain unclear, while dexamethasone is found to be a double-edged sword during breast cancer progression and metastasis [16]. At lower concentrations, Dexamethasone inhibits the growth and metastasis of breast cancer tumors, while higher concentrations may promotes breast cancer progression.

To explore candidate drugs by molecular docking simulation *PCK1* and *LPL* are utilized as the target receptors to analyze their interactions with the chemicals (bis(4-hydroxyphenyl)sulfone, Dexamethasone, Triflumizole, and Rosiglitazone). The molecular analysis results show that bis(4-hydroxyphenyl)sulfone, Dexamethasone, Triflumizole, and Rosiglitazone have high binding affinity with the target receptors. These chemicals could be essential for discovering potential drug candidates for the treatment of recurrent breast cancer metastasis to the liver and should be further subjected to testing.

Our study has some limitations since it relies solely on bioinformatics analysis and lacks experimental verification; more clinical trials and researches are needed to validate the miRNAs, TFs, chemical compounds, and potential drugs. Despite using algorithm-based scientific methodologies, our research serves to predict the potential biomarkers and therapeutic targets for treating recurrent breast cancer metastasis to the liver.

4. Materials and Methods

4.1. Identification of Datasets and Analysis of Differentially Expressed Genes

In this study, a combined bioinformatics approach was used to find essential differentially expressed genes (DEGs) and describe the functions of these genes by combining different data sets. Gene expression datasets related to advanced local and liver metastasis breast cancer were obtained from the National Center for Biotechnology Information's Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) (accessed on 9 November 2023) [17]. Local and liver metastasis samples were retrieved from the GSE46141 (n=30) and GSE56493 (n=43) datasets. Their respective metastatic recurrence information was retrieved from the GSE175692 dataset. Differential gene expression (DEG) analysis between the following groups was performed: (1) breast cancer local versus liver metastasis (2) breast cancer metastasis in breast anatomical site versus breast cancer metastasis in liver anatomical site (3) relapsed metastatic breast cancer in breast organ versus liver. We used the following cut-off threshold values to identify DEGs: log2(fold change [FC]) value>2, p-value<0.05, and Bonferroni false discovery rate (FDR) < 0.05. To identify the specific and commonly dysregulated genes among the DEGs identified in each of the datasets, we performed a Venn diagram analysis.

4.2. Clustering and Analysis of Identified DEGs

To classify the analyzed samples based on gene expression profiles and to observe the overall gene expression patterns in each condition of the datasets, we performed hierarchical clustering using the Cluster, heatmap option in the SR plot (https://www.bioinformatics.com.cn/srplot) (accessed on 26 November 2023) which employ *pheatmap* R package [18]. To compare the expression of identified genes in each of the breast cancer types and the normal at a threshold of p-value < 0.05, we used GEPIA 2.0 (http://gepia2.cancer-pku.cn/) [19], a web-based tool that utilizes data from the TCGA and Genotype-Tissue Expression (GTEx) databases [20] (accessed on 9 December 2023). To analyze the prognostic value of DEGs, we analyzed the Copy number variants (CNV) followed by survival analysis using the Breast cancer integrated platform (http://www.omicsnet.org/bcancer/) [21] using the default settings. (accessed on 13 January 2024)

4.3. Identification of Genetic Alterations in the DEGs

To identify the genetic alterations in the DEGs, we used cBioprtal (https://www.cbioportal.org/) [22] (accessed on 18 January 2024). We selected the following breast datasets in cBioPortal for analyzing the genomic alterations in the identified DEGs: Metastatic Breast Cancer (INSERM, PLoS Med 2016); Metastatic Breast Cancer (MSK, Cancer Discovery 2022); The Metastatic Breast Cancer Project (Archived, 2020); The Metastatic Breast Cancer Project (Provisional, December 2021); Breast Cancer (HTAN, 2022); Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016); Breast Cancer (MSK, Clinical Cancer Res 2020); Breast Cancer (MSK, NPJ Breast Cancer 2019); Breast Cancer (SMC 2018); Breast Cancer Xenografts (British Columbia, Nature 2015); Breast Invasive Carcinoma (British Columbia, Nature 2012); Breast Invasive Carcinoma (Broad, Nature 2012); Breast Invasive Carcinoma (Sanger, Nature 2012); Breast Invasive Carcinoma (TCGA, Cell 2015); Juvenile Papillomatosis and Breast Cancer (MSK,J Pathol. 2015); MAPK on resistance to anti-HER2 therapy for breast cancer (MSK, Nat Commun. 2022).

4.4. Immune Cell Infiltration Analysis of DEGs

To explore correlations between the DEGs and identify the clinical relevance of tumor-immune infiltrations we used the TIMER 2.0 (http://cistrome.org/TIMER/) [23] database (accessed on 8 January 2024). Using TIMER 2.0, we conducted univariate and multivariate Cox regression analyses using variables such as tumor purity, tumor stage and hub gene expression.

4.5. Identify Positive Correlated Genes

To determine whether a statistically significant linear relationship existed between the DEGs and other potential genes, we use UALCAN(https://ualcan.path.uab.edu/) (accessed on 2 January 2024) [24]. The tool utilizes TCGA datasets to measure the Pearson correlation between genes given

their expression level in breast invasive carcinoma. If a coefficient value lied between 0.4 and 1, then there is a medium to strong correlation exists.

4.6. Gene to miRNA Interaction

To identify miRNA regulatory network for our hub genes in Human, we used miRNet 2.0 (https://www.mirnet.ca) /)(accessed on 3 February 2024) [25], a platform that uses 14 different database (TarBase, miRTarBase, miRecords, miRanda- S mansoni only, miR2Disease, HMDD, PhenomiR, SM2miR, PharmacomiR, EpimiR, starBase, TransmiR, ADmiRE, and TAM 2.0) and uses miRanda and TarPmiR to predict genomic targets for microRNAs.

4.7. NetworkAnalysis on Hub Genes

We used the NetworkAnalyst (https://www.networkanalyst.ca/) (accessed on 8 February 2024) [26] online tool to visualize a hub gene–transcription factor (gene-TF) interaction network, and protein–chemical (gene-chemical) interaction. To construct a gene-TF network, the TF and gene target data were derived from TRRUST [27], a curated database of human transcriptional regulatory networks. To create gene–chemical interaction network, we used the Comparative Toxicogenomics Database [28], a database that provide information about chemical–gene/protein interactions, chemical–disease and gene–disease relationships.

4.8. Construction of Gene Networks and Protein-Protein Interactions

To identify the functionally related genes to our query genes and to perform protein-protein interaction analysis, we used the GeneMANIA (https://genemania.org/)(accessed on 22 December 2023) [29] prediction server, STRING online analysis tool (http://www.string-db.org/)/)(accessed on 22 December 2023) [30] and CytoHubba plugin in Cystoscape [31]. In GeneMANIA, we selected Homo sapiens as the target organism and interaction score > 0.4 for our analysis. In CytoHubba we selected the top 10 genes ranked by degree for our analysis. To identify the signaling relationship associated with the up and down-regulation of the genes, we used Signor 2.0 (https://signor.uniroma2.it/APIs.php) [32]. (accessed on 15 December 2023)

4.9. Identification of Potential Treatment Targets

To identify potential treatment targets, the DEGs were analyzed using the Drug Signature Database (DSigDB), sourced from the *Enrichr* platform (https://amp.pharm.mssm.edu/Enrichr/)(accessed on 12 February 2024) [33]. Enrichr provided detailed visualizations of the common functions associated with the input genes, exploring drug molecules associated with the DEGs.

4.10. Analyzing the Unique Ligands and Their Respective Binding to DEGs

We accessed the RCSB Protein Data Bank (PDB) (https://www.rcsb.org) [34] to obtain the three dimensional structure of *PCK1* (PDB ID: 1NHX) and *LPL* (PDB ID: 6E7K). To prepare the protein, we used Discovery Studio Visualizer 3.0 (https://discover.3ds.com/discovery-studio-visualizer-download) /) [35] (accessed on 24 March 2024) [36] to remove the water and heteroatoms compounds, and add polar hydrogens to the protein structure and save as PDB format. Then we used PubChem (https://pubchem.ncbi.nlm.nih.gov/)(accessed on 24 March 2024) [36] to download the structure of the chemical compounds [Triflumizole (Pubchem ID: 91699), bis(4-hydroxyphenyl)sulfone (Pubchem ID: 6626), Dexamethasone (Pubchem ID: 5743) and rosiglitazone (Pubchem ID: 77999)] and use SwissADME(http://www.swissadme.ch/index.php) (accessed on 25 March 2024) [37] to obtain the pharmacokinetics, drug-likeness and medicinal chemistry friendliness information of the chemicals. The selected chemical compounds and the newly modified protein are provided as input into the Virtual Screening software interface PyRx (https://pyrx.sourceforge.io/)(accessed on 28 March 2024) [38] for docking studies. To analyze the result from the PyRx interface, we enter the relevant docking

output files back to Discovery Studio Visualizer 3.0 to determine the interaction between the respective ligands and the receptors.

4.11. Gene Ontology (GO) and Functional Enrichment Analysis

Functional annotation and pathway enrichment analyses of the genes were performed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID; v6.8, http://david.ncifcrf.gov) database (accessed on 26 November 2023) [39]. Gene Ontology (GO) term enrichment was performed using the enrichment bubble option in the SR plot (https://www.bioinformatics.com.cn/srplot) (accessed on 10 November 2023) [40] which uses the ggplot2 [41] function in the R package.

5. Conclusions

In conclusion, this study utilized an integrative bioinformatics approach to uncover critical genes and pathways involved in breast cancer liver metastasis. By comparing gene expression profiles across multiple datasets, 5 robust and consistent differentially expressed genes (DEGs) were identified between primary tumors and liver metastases. Further pathway and network analysis on these DEGs and correlated interacting partners revealed the enrichment of metabolic processes related to lipid and glucose metabolism. In particular, the genes *PCK1* and *LPL* emerge as central hubs exhibiting altered expression.

References

1. Chen, M., W. Zheng, and L. Fang, *Identifying liver metastasis-related hub genes in breast cancer and characterizing.* PeerJ, 2023. **11**: p. e15311.
2. Rashid, N.S., et al., *Breast cancer liver metastasis: current and future treatment approaches.* Clin Exp Metastasis, 2021. **38**(3): p. 263-277.
3. Zhao, H.Y., et al., *Incidence and prognostic factors of patients with synchronous liver metastases upon initial diagnosis of breast cancer: a population-based study.* Cancer Manag Res, 2018. **10**: p. 5937-5950.
4. Golestan, A., et al., *Unveiling promising breast cancer biomarkers: an integrative approach combining bioinformatics analysis and experimental verification.* BMC Cancer, 2024. **24**(1): p. 155.

5. Nguyen, T.N.Q., et al., *The regulation of insulin receptor/insulin-like growth factor 1 receptor ratio, an important factor for breast cancer prognosis, by TRIP-Br1.* J Hematol Oncol, 2022. **15**(1): p. 82.

6. Xiang, J., K. Wang, and N. Tang, *PCK1 dysregulation in cancer: Metabolic reprogramming, oncogenic activation, and therapeutic opportunities.* Genes Dis, 2023. **10**(1): p. 101-112.

7. Thomas, G.W.C., et al., *Origins and Long-Term Patterns of Copy-Number Variation in Rhesus Macaques.* Mol Biol Evol, 2021. **38**(4): p. 1460-1471.

8. Jiang, H., et al., *A newly discovered role of metabolic enzyme PCK1 as a protein kinase to promote cancer lipogenesis.* Cancer Commun (Lond), 2020. **40**(9): p. 389-394.

9. Tobin, A.J., et al., *Lipoprotein lipase hydrolysis products induce pro-inflammatory cytokine expression in triple-negative breast cancer cells.* BMC Res Notes, 2021. **14**(1): p. 315.

10. Ahmad, M., *In-Silico Target Prediction for hsa-mir-27a and Identification of Genes Involved in Breast Cancer*, A.A. Shah, Editor. 2019: International journal of human genetics. p. 12-21.

11. Elango, R., et al., *MicroRNA Expression Profiling on Paired Primary and Lymph Node Metastatic Breast Cancer Revealed Distinct microRNA Profile Associated With LNM.* Front Oncol, 2020. **10**: p. 756.

12. Korpal, M. and Y. Kang, *The emerging role of miR-200 family of microRNAs in epithelial-mesenchymal transition and cancer metastasis.* RNA Biol, 2008. **5**(3): p. 115-9.

13. Liu, G.E., et al., *Identification of conserved regulatory elements in mammalian promoter regions: a case study using the PCK1 promoter.* Genomics Proteomics Bioinformatics, 2008. **6**(3-4): p. 129-43.

14. Gao, Y., et al., *SP1 Expression and the Clinicopathological Features of Tumors: A Meta-Analysis and Bioinformatics Analysis.* Pathol Oncol Res, 2021. **27**: p. 581998.

15. Keledjian, K., et al., *The peroxisome proliferator-activated receptor gamma (PPARγ) agonist, rosiglitazone, ameliorates neurofunctional and neuroinflammatory abnormalities in a rat model of Gulf War Illness.* PLoS One, 2020. **15**(11): p. e0242427.

16. Pang, J.M., et al., *Effects of synthetic glucocorticoids on breast cancer progression.* Steroids, 2020. **164**: p. 108738.

17. Edgar, R., M. Domrachev, and A.E. Lash, *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.* Nucleic Acids Res, 2002. **30**(1): p. 207-10.

18. R., K., *pheatmap: Pretty Heatmaps. .* 2019.

19. Tang, Z., et al., *GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis.* Nucleic Acids Res, 2019. **47**(W1): p. W556-w560.

20. *Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans.* Science, 2015. **348**(6235): p. 648-60.

21. Wu, J., et al., *BCIP: a gene-centered platform for identifying potential regulatory genes in breast cancer.* Sci Rep, 2017. **7**: p. 45235.

22. Cerami, E., et al., *The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.* Cancer Discov, 2012. **2**(5): p. 401-4.

23. Li, T., et al., *TIMER2.0 for analysis of tumor-infiltrating immune cells.* Nucleic Acids Res, 2020. **48**(W1): p. W509-w514.

24. Chandrashekar, D.S., et al., *UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses.* Neoplasia, 2017. **19**(8): p. 649-658.

25. Chang, L., et al., *miRNet 2.0: network-based visual analytics for miRNA functional analysis and systems biology.* Nucleic Acids Res, 2020. **48**(W1): p. W244-w251.

26. Zhou, G., et al., *NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis.* Nucleic Acids Res, 2019. **47**(W1): p. W234-w241.

27. Han, H., et al., *TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions.* Nucleic Acids Res, 2018. **46**(D1): p. D380-d386.

28. Davis, A.P., et al., *The Comparative Toxicogenomics Database: update 2017.* Nucleic Acids Res, 2017. **45**(D1): p. D972-d978.

29. Mostafavi, S., et al., *GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function.* Genome Biol, 2008. **9 Suppl 1**(Suppl 1): p. S4.

30. Szklarczyk, D., et al., *The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets.* Nucleic Acids Res, 2021. **49**(D1): p. D605-d612.

31. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.

32. Licata, L., et al., *SIGNOR 2.0, the SIGnaling Network Open Resource 2.0: 2019 update.* Nucleic Acids Res, 2020. **48**(D1): p. D504-d510.

33. Kuleshov, M.V., et al., *Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.* Nucleic Acids Res, 2016. **44**(W1): p. W90-7.

34. Goodsell, D.S. and S.K. Burley, *RCSB Protein Data Bank tools for 3D structure-guided cancer research: human papillomavirus (HPV) case study.* Oncogene, 2020. **39**(43): p. 6623-6632.

35. *BIOVIA DS. BIOVIA discovery studio visualizer. 2016. 20:779.*

16

36. Kim, S., et al., *PubChem 2019 update: improved access to chemical data.* Nucleic Acids Res, 2019. **47**(D1): p. D1102-D1109.

37. Daina, A., O. Michielin, and V. Zoete, *SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules.* Sci Rep, 2017. **7**: p. 42717.

38. Dallakyan, S. and A.J. Olson, *Small-molecule library screening by docking with PyRx.* Methods Mol Biol, 2015. **1263**: p. 243-50.

39. Dennis, G., Jr., et al., *DAVID: Database for Annotation, Visualization, and Integrated Discovery.* Genome Biol, 2003. **4**(5): p. P3.

40. Tang, D., et al., *SRplot: A free online platform for data visualization and graphing.* PLoS One, 2023. **18**(11): p. e0294236.

41. H., W., *ggplot2: Elegant Graphics for Data Analysis. Springer.* 2016.