Saudi Journal of Biological Sciences 29 (2022) 103318

Contents lists available at ScienceDirect

Saudi Journal of Biological Sciences

journal homepage: www.sciencedirect.com

Original article

Comprehensive analysis of gene expression profiles to identify differential prognostic factors of primary and metastatic breast cancer

Sarah Albogami

Department of Biotechnology, College of Science, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia

ARTICLE INFO

Article history:
Received 7 February 2022
Revised 17 April 2022
Accepted 19 May 2022
Available online 23 May 2022

Keywords:
Breast cancer
Metastasis
Prognostic marker
Overall survival
Differentially expressed genes
Protein-protein interaction

ABSTRACT

Breast cancer accounts for nearly half of all cancer-related deaths in women worldwide. However, the molecular mechanisms that lead to tumour development and progression remain poorly understood and there is a need to identify candidate genes associated with primary and metastatic breast cancer progression and prognosis. In this study, candidate genes associated with prognosis of primary and metastatic breast cancer were explored through a novel bioinformatics approach. Primary and metastatic breast cancer tissues and adjacent normal breast tissues were evaluated to identify biomarkers characteristic of primary and metastatic breast cancer. The Cancer Genome Atlas-breast invasive carcinoma (TCGA-BRCA) dataset (ID: HS-01619) was downloaded using the mRNASeq platform. Genevestigator 8.3.2 was used to analyse TCGA-BRCA gene expression profiles between the sample groups and identify the differentially-expressed genes (DEGs) in each group. For each group, Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway enrichment analyses were used to determine the function of DEGs. Networks of protein–protein interactions were constructed to identify the top hub genes with the highest degree of interaction. Additionally, the top hub genes were validated based on overall survival and immunohistochemistry using The Human Protein Atlas. Of the top 20 hub genes identified, four (*KRT14*, *KIT*, *RAD51*, and *TTK*) were considered as prognostic risk factors based on overall survival. *KRT14* and *KIT* expression levels were upregulated while those of *RAD51* and *TTK* were downregulated in patients with breast cancer. The four proposed candidate hub genes might aid in further understanding the molecular changes that distinguish primary breast tumours from metastatic tumours as well as help in developing novel therapeutics. Furthermore, they may serve as effective prognostic risk markers based on the strong correlation between their expression and patient overall survival.

1. Introduction

Breast cancer (BC), which is the most frequently diagnosed cancer type in women, could be treatable if detected in its early stages (Andreopoulou and Hortobagyi, 2008, Jemal et al., 2009, Ozkan, 2022, Masood, 2005, Albogami et al., 2021). Based on its stage, the prognosis of BC is usually classed from I to IV, with various sub-stages(Cianfrocca and Goldstein, 2004). BC is considered malignant when it progresses from stage I and II to advanced disease (stage III), and subsequently metastasizes to the lungs, bone, brain, and other organs (stage IV) (Linde et al., 2018). Numerous aspects of BC can aid in predicting the risk of recurrence and the likelihood of survival following surgery (Maitra and Srivastava, 2022). The traditional prognostic factors of BC include positive axillary nodes, tumour size, histology, grade, lymphovascular invasion, the expression of the hormone receptor, and HER2 status (Ly et al., 2012). These factors have been discussed in detail, with an emphasis on classification issues (Eliyatkın et al., 2015, Kim et al., 2012, Sauerbrei et al., 1999). Modern molecular prognostic markers might be able to provide information beyond that by traditional prognostic,factors and this was achieved by a better comprehension of BC pathophysiology, which resulted in the

Abbreviations: BC, breast cancer; BP, biological process; CC, cellular component; CI, confidence interval; DEG, differentially expressed gene; FDR, false discovery rate; GEPIA, gene expression profiling interactive analysis; GO, gene ontology; HR, hazard ratio; IDC, infiltrating ductal carcinoma; KEGG, Kyoto Encyclopedia of Genes and Genomes; MCODE, molecular complex detection; MF, molecular function; OS, overall survival; PPI, protein-protein interaction; RNA-Seq, RNA sequencing; STRING, search tool for the retrieval of interacting genes; TCGA-BRCA, The Cancer Genome Atlas-breast invasive carcinoma.

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

E-mail address: dr.sarah@tu.edu.sa

https://doi.org/10.1016/j.sjbs.2022.103318
1319-562X/

development of novel molecular markers, enhanced risk evaluation, improved therapies, and personalised treatment (Faramarzi et al., 2021, Cilibrasi et al., 2021, Zhou et al., 2022). Genetic profiling can yield predictive and prognostic gene expression signatures, which might aid in the classification of cancers and provide new targeted therapy (Loughman et al., 2022). The focus on new prognostic indicators arises from the belief that a considerable proportion of patients with BC in the early stages are not diagnosed by accurate methods, usually using microscopic metastases (Rogers et al., 2002).

To date, only a few effective prognostic and predictive indicators have been used clinically to manage patients with BC (Kumar et al., 2012). Although numerous techniques have been developed over the years, none have proven to be clinically useful. With the improvement in next-generation sequencing techniques, RNA sequencing (RNA-Seq) technology has emerged as a novel, straightforward, and helpful tool for assessing the content of cDNA transcripts and analysing differential gene expression utilizing high-throughput sequencing (Negi et al., 2022). RNA-Seq technology has demonstrated a high sensitivity for biomarker discovery, while increasing the resolution and decreasing the associated time and cost (Stark et al., 2019, Ergin et al., 2022, Hwang et al., 2018). Thus, cancer-associated genes and the associated signalling pathways at the genome level can be identified using RNA-Seq and other high-throughput sequencing technologies (Ren et al., 2012, Zhang et al., 2019a).

RNA-Seq technology is not frequently used by molecular diagnostic laboratories for clinical testing, monitoring, or management of patients with BC (Roychowdhury and Chinnaiyan, 2016). Owing to this, a sizable fraction of potentially diagnosable cases could be unresolved at the moment (Marco-Puche et al., 2019). Certain authors have illustrated the value of RNA sequencing in diagnosing 10% of patients and identifying potential genes for the other 90% (Kremer et al., 2017). However, the application of RNA-Seq in clinical settings is severely limited owing to the generation of large datasets that require skilled analysis, the inability to maintain independence, poor validation, lack of methodological standardisation, and the complexity of RNA-Seq data outcomes.(Salto-Tellez and Gonzalez de Castro, 2014, Lightbody et al., 2019, Nazarov et al., 2017). Therefore, further studies are required to ensure that RNA-Seq is consistently and reliably performed in clinics. Although RNA-Seq has been used for cancer transcriptome profiling, limited studies have been conducted to identify differences in the transcriptomes of normal breast, primary BC, and metastatic BC tissues using RNA-Seq. Hence, a technique that is applicable to routine clinical practice must be developed based on differential gene profiling. This is especially critical for the identification of high-risk cancer cells that may undergo metastases (Feng et al., 2007).

Therefore, by gene expression profiling and comprehensive bioinformatics analysis of The Cancer Genome Atlas-breast invasive carcinoma (TCGA-BRCA) dataset, this study aimed to compare the primary and metastatic BC tissues to normal breast tissue to screen for the top 20 hub genes, annotate differentially-expressed gene (DEG) profiles, integrate the gene expression profiles with clinical data, and finally identify biomarkers involved in primary and metastatic BC.

2. Materials and methods

2.1. BC transcriptome and clinical profile acquisition and workflow

Human BC transcriptome and clinical profiles [Project title: TCGA-BRCA; Experimenter: NIH national cancer institute (Koboldt et al., 2012, Ciriello et al., 2015); Experiment ID: HS-01619; Experimental technology platform: HS_mRNASeq_HUMAN_GL: mRNA-Seq Gene Level Homo sapiens (ref: Ensembl 97, GRCh38.p12); Study design: DEGs between primary IDC and tumour adjacent tissue, and DEGs between metastatic IDC and tumour adjacent tissue] were analysed and downloaded from Genevestigator 8.1.0 (Zurich, Switzerland) in May 2021. Supplementary Table S1 contains the demographic information for the datasets, and Fig. 1 depicts the study workflow.

2.2. Identification of DEGs between groups

DEGs were analysed in two groups: primary tumour tissue samples obtained from patients with infiltrating duct carcinoma compared with normal breast tissue samples (P vs. N) and metastatic tumour tissue samples obtained from axillary lymph nodes of patients with primary infiltrating duct carcinoma of the breast compared with normal breast tissue samples (M vs. N). DEGs were considered upregulated and downregulated if |Log2 fold change| $\geq$ 2 and |Log2 fold change| ≥ -2, respectively, and p-value < 0.05 were found between sample groups, in accordance with previously specified criteria (Ward et al., 2013, Luo et al., 2017). Genevestigator 8.3.2 was used to analyse TCGA-BRCA gene expression profiles between sample groups.

2.3. Protein–protein interaction (PPI) network construction, hub gene identification, and module analysis

PPI networks were constructed between DEGs in the P vs. N and M vs. N groups using the search tool for the retrieval of interacting genes (STRING) database (Jensen et al., 2009). The STRING database was used to obtain statistical data for each PPI network. The Network Analyser tool of the Cytoscape software (version: 3.8.2) with indirect parameters was employed to visualise as well as analyse the PPI networks to determine the interactional correlations among DEGs (Shannon et al., 2003). The top 20 hub genes from each PPI network group were chosen due to their high degree of connectivity ($\geq$10). Additionally, the Cytoscape plugin molecular complex detection (MCODE) was utilized to identify clusters of genes that were highly connected across the entire PPI network.

2.4. Functional enrichment analysis

The inputs of PPI networks and networks for the top twenty candidate hub genes for both P vs. N and M vs. N groups were applied to find the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and gene ontology (GO) terms. The ClueGo and Cluepedia Cytoscape plugins were used for GO enrichment analyses (Bindea et al., 2013). GO annotations describe the relationship that exists between gene expression and molecular function (MF), cellular component (CC), and biological process (BP).

2.5. Validation of hub genes

Gene expression profiling interactive analysis (GEPIA), a web server, was applied to calculate the Kaplan–Meier curve by correlating gene expression with time and the percent of overall survival (OS) for the top 20 hub gene candidates identified in both P vs. N and M vs. N groups to determine whether they have an effect on survival of patients with BC (Tang et al., 2017). Significant differences were defined as those with a p value of < 0.05, and the median was employed as the group cut-off; high and low cut-offs were 50%. The hazard ratio (HR) with a 95% confidence interval was determined using the Cox PH model, and the BRCA dataset was used as the selection dataset. The Human Protein Atlas (www.proteinatlas.org) database was used to obtain immunohistochemistry

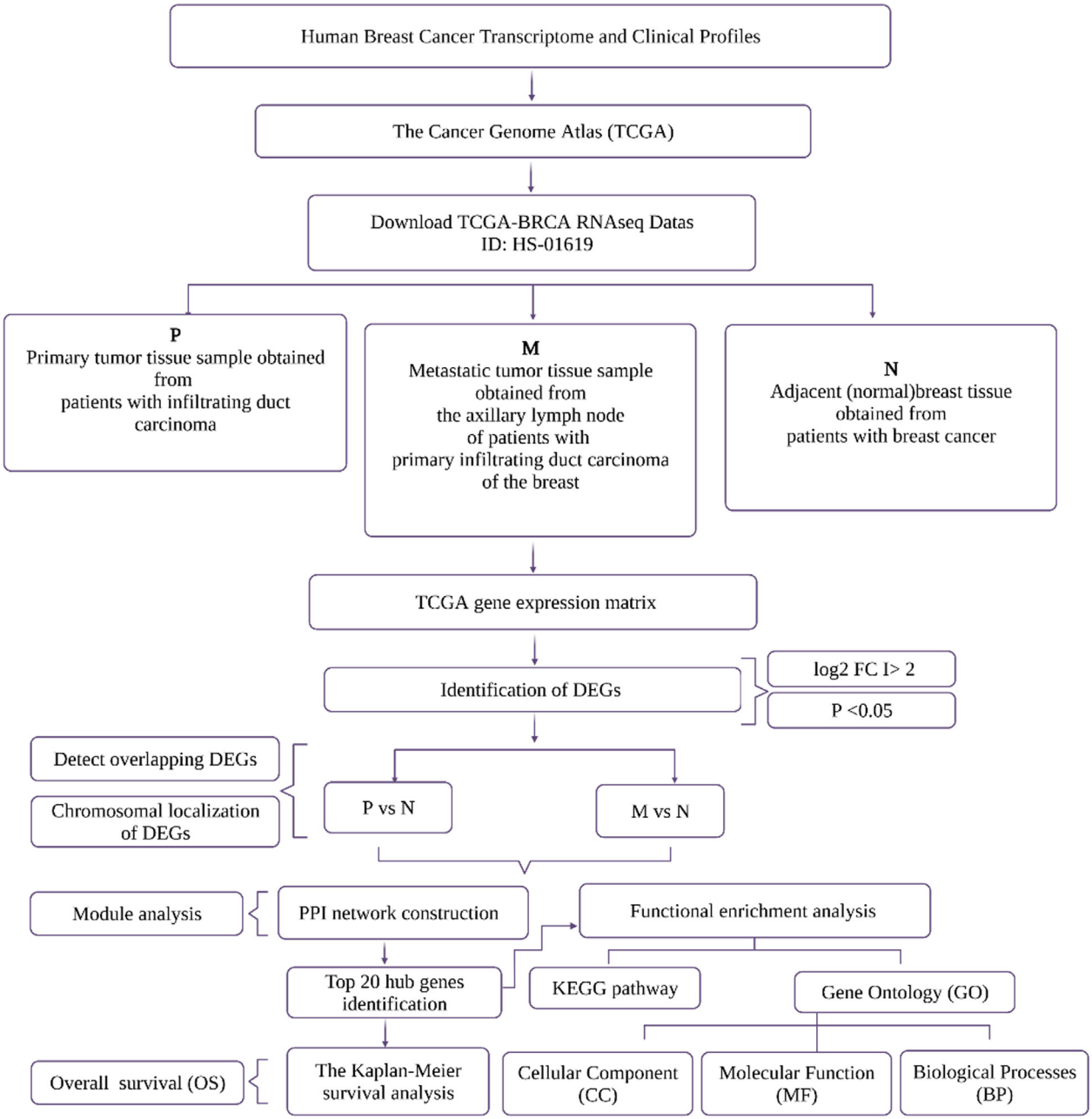

Fig. 1. Flowchart depicting the study design and steps involved in the processing, analysis, and validation of data.

results for the top 20 hub genes in each group; this allowed us to compare the dysregulation of specific hub genes across groups.

2.6. Statistical analysis

GraphPad Prism ver. 9.1.2 (San Diego, CA, USA) was utilized to create a volcano plot to reveal upregulated and downregulated DEGs. The data are expressed as the mean ± SD. The fold-change (FC) was determined using the log2 value of the signal intensity; FC > 1.5 was deemed significant. At a $p < 0.05$, the differences in gene expression between the datasets were declared significant. Venn Painter 1.2.0 (Toronto, Canada) was used to illustrate the overlap in DEGs between sample groups (Lin et al., 2016). DEGs were mapped to chromosomes using Circa (CA, USA). The RStudio, PBC software (version: 1.3.1056) was used to identify the KEGG pathways in both groups for the complete set of DEGs and the top twenty hub genes. The ggplot2 and tidyr packages in R were used to visualise the KEGG enrichment analysis results. The inputs for KEGG enrichment analysis were generated from the DAVID bioinformatics resource 6.8 online database (Sherman and Lempicki, 2009, Huang et al., 2009).

3. Results

3.1. Analysis revealed 528 and 955 DEGs in P vs. N and M vs. N groups, respectively

In total, 528 DEGs (134 upregulated and 394 downregulated) were identified between primary tumour tissues and normal breast tissues (P vs. N group) (Fig. 2a, Supplementary Table S2), and 955 DEGs (422 upregulated and 533 downregulated) were identified between metastatic tumour tissue and normal breast tis-

sue from patients with BC (M vs. N group) (Fig. 2b, Supplementary Table S2). DEGs in both groups were localised on chromosomes (Fig. 2c, d), and a Venn diagram revealed 99 overlapping upregulated and 201 overlapping downregulated DEGs between the P vs. N and M vs. N groups (Fig. 3a, b).

Among the genes that were not common between the groups, 35 genes were upregulated and 193 genes were downregulated only in the P vs. N group (Fig. 3a, b), while 323 genes were upregulated and 332 were downregulated only in the M vs. N group (Fig. 3a, b). The top 10 unique DEGs (upregulated and downregulated) that did not overlap between the two groups are summarised in Supplementary Tables S2 (P vs. N) and S3 (M vs. N). Supplementary Tables S4 and S5 present the results of the Venn diagram for all DEGs.

3.2. PPI network construction and analysis revealed genes potentially involved in early metastasis

The STRING database was employed to construct the PPI network and analyse its features. The P vs. N PPI network consisted of 219 nodes connecting 679 edges, with an average neighbour count of 7.564, average clustering coefficient of 0.387, network heterogeneity of 1.122, 39 connected components, and a network centralisation of 0.221 (Fig. 4a). In contrast, the M vs. N PPI network consisted of 623 nodes connecting 2120 edges, with an average neighbour count of 7.895, average clustering coefficient of 0.248, network heterogeneity of 1.304, 81 connected components, and network centralisation of 0.096 (Fig. 4b).

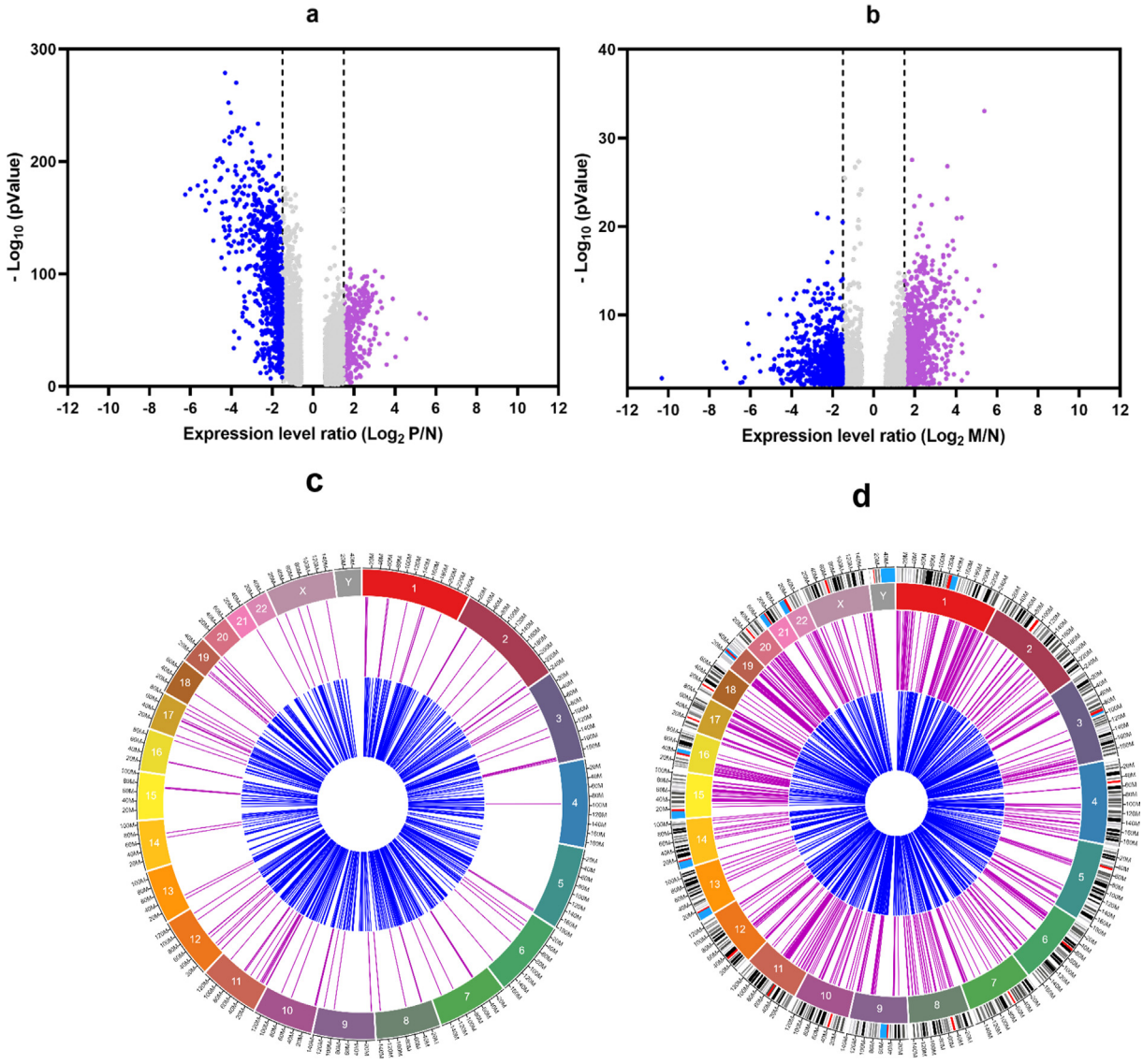

Fig. 2. Differentially expressed genes (DEGs) between P vs. N and M vs. N groups. (**a**, **b**) Volcano plots of DEGs, between a) P vs. N. and b) M vs. N groups. Volcano plots present upregulated expression with purple colour, and downregulated expression with blue colour. (**c**, **d**) CIRCOS circular plot of DEGs in the genome. (c) Upregulated and downregulated DEGs in P vs. N group. (**b, d**) Upregulated and downregulated DEGs in M vs. N group. CIRCOS circular plot presents the chromosome number with track 1, DEGs with upregulated expression is shown in purple colour with track 2, and DEGs with downregulated expression is shown with track 3. DEGs were considered upregulated and downregulated if a |Log2 fold change| ≥ 2 and |Log2 fold change| ≥ -2, respectively, and p-value < 0.05. P vs. N, primary tumour tissue samples obtained from patients with infiltrating duct carcinoma compared with normal breast tissue samples; M vs. N, metastatic tumour tissue samples obtained from axillary lymph nodes of patients with primary infiltrating duct carcinoma of the breast compared with normal breast tissue samples.

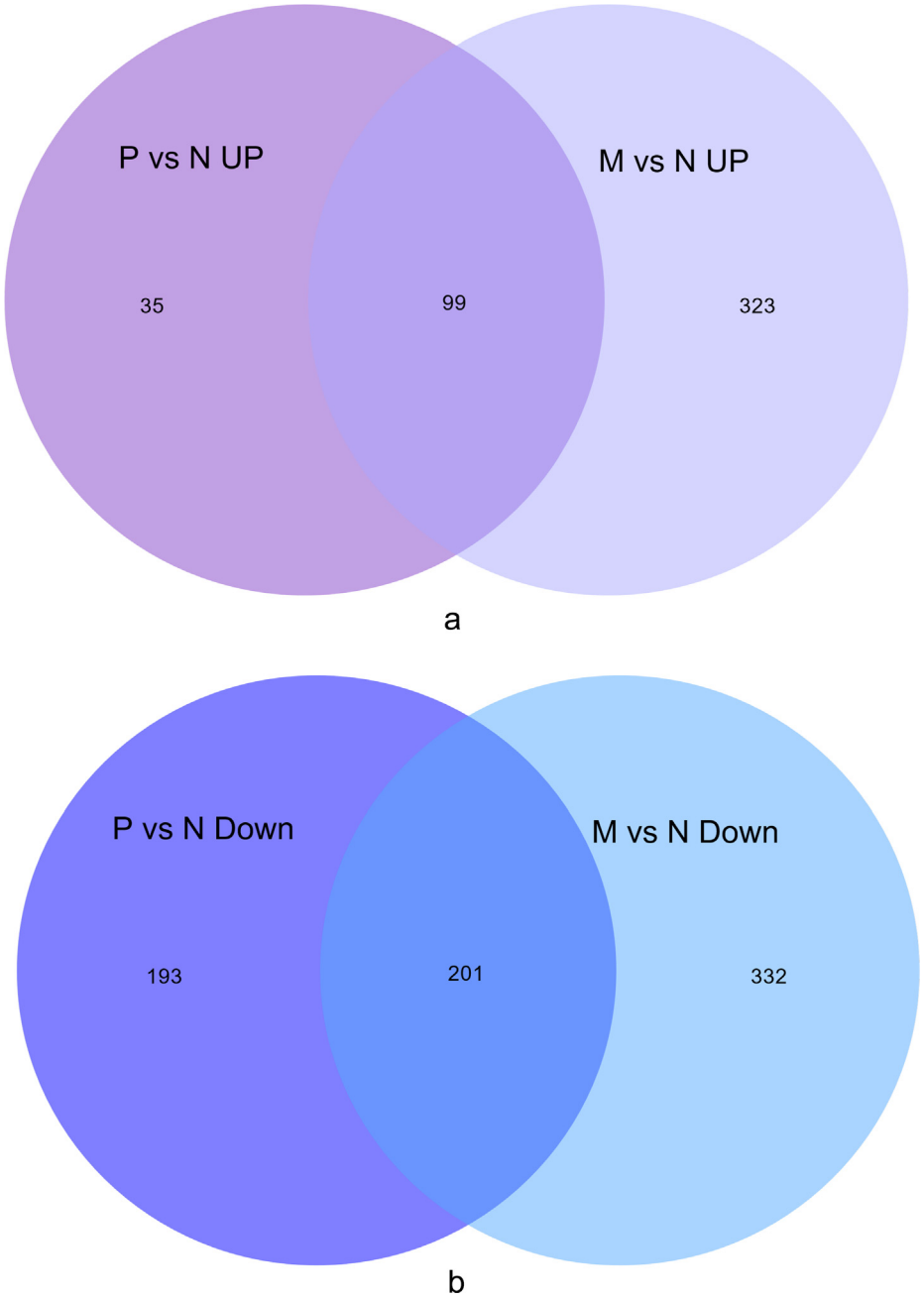

Fig. 3. Venn diagram of differentially expressed genes (DEGs) between P vs. N and M vs. N groups. (**a**) Number of upregulated DEGs between P vs. N and M vs. N groups. (**b**) Number of downregulated DEGs between P vs. N and M vs. N groups. Each group (P vs. N and M vs. N) is represented in an independent circle; the overlapping area between the circles indicates the overlap of DEGs between the groups, and the number in each circle indicates the unique DEGs. Purple indicates upregulated DEGs, and blue indicates downregulated DEGs. Venn Painter 1.2.0 was utilized to illustrate the overlap in DEGs between sample groups. P vs. N, primary tumour tissue samples obtained from patients with infiltrating duct carcinoma compared with normal breast tissue samples; M vs. N, metastatic tumour tissue samples obtained from axillary lymph nodes of patients with primary infiltrating duct carcinoma of the breast compared with normal breast tissue samples.

With degree $\geq$ 10 as the cut-off criterion, 47 and 112 PPI network hub genes were identified among the P vs. N (Supplementary Table S6) and M vs. N (Supplementary Table S7) DEGs, respectively. The top 20 genes in P vs. N and M vs. N groups are listed in Tables 1 and 2, respectively. These hub genes may be involved in early metastasis.

When the cut-off criterion of MCODE was defined as $\geq$ 5, only one cluster was identified from the P vs. N DEG-PPI network (Fig. 5a), comprising 21 nodes and 95 edges, and 12 seed genes that were common with the top 20 hub genes (*DGAT2*, *FABP4*, *PCK1*, *ADIPOQ*, *COL1A1*, *SPP1*, *LPL*, *LIPE*, *LEP*, *FOS*, *FN1*, and *PLIN1*). Furthermore, five clusters were identified in the M vs. N DEG-PPI network (Fig. 5b–f), and the most significant cluster (score 29.429) consisted of 36 nodes and 515 edges, with 15 seed genes shared with the top 20 hub genes (*TTK*, *TYMS*, *OIP5*, *MAD2L1*, *RACGAP1*, *NDC80*, *RAD51*, *MCM10*, *KIF15*, *KIF11*, *CDC25C*, *ESPL1*, *CCNA2*, *PLK4*, and *RAD51AP1*; Fig. 5b). The second cluster with a score of 8.000 consisted of 12 nodes and 44 edges and contained three seed genes shared with the top 20 hub genes (*CCND1*, *MYC*, and *BMP4*; Fig. 5c). The third (11 nodes and 25 edges), fourth (15 nodes and 35 edges), and fifth (5 nodes and 10 edges) clusters with a score of 5.000 each did not share any seed genes with the top 20 hub

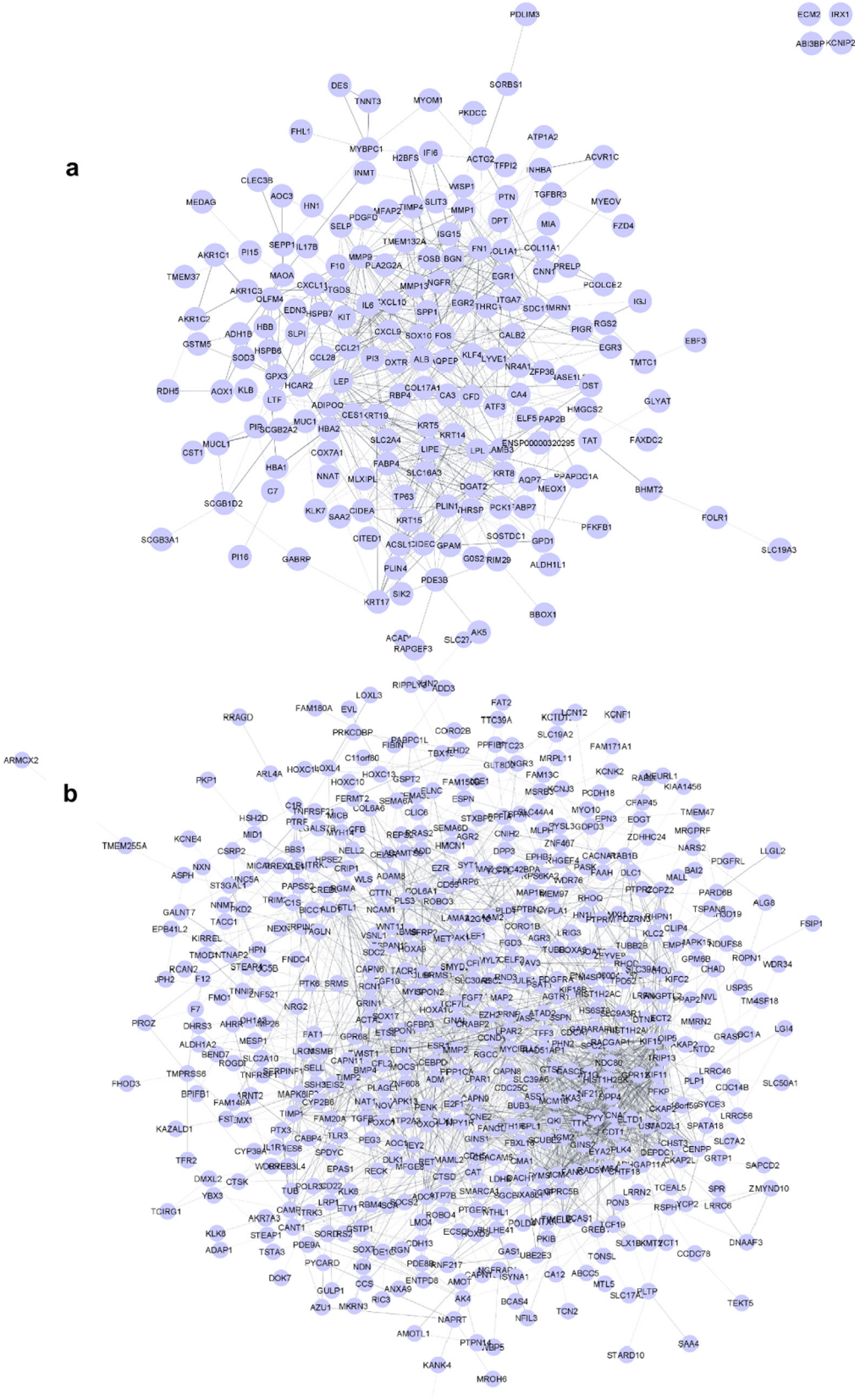

Fig. 4. A diagrammatic representation of the protein–protein interaction (PPI) network. (**a**) P vs. N PPI network. (**b**) M vs. N PPI network. The Cytoscape software was used to create the figures. The STRING database was used to obtain statistical data for each PPI network. The Network Analyser tool of the Cytoscape software with indirect parameters was employed to visualise as well as analyse the PPI networks to determine the interactional correlations of DEGs. P vs. N, primary tumour tissue samples obtained from patients with infiltrating duct carcinoma compared with normal breast tissue samples; M vs. N, metastatic tumour tissue samples obtained from axillary lymph nodes of patients with primary infiltrating duct carcinoma of the breast compared with normal breast tissue samples.

Table 1
Top 20 hub genes in P vs. N group from the PPI network.

ID	Gene symbol	Description	Log-fold change	P-value	FDR	Degree of connectivity	Up or down
ENSG00000163631	*ALB*	Albumin	−2.51152	1.46129E-49	2.80903E-48	57	↓
ENSG00000115414	*FN1*	Fibronectin 1	3.051057	1.69557E-47	3.03701E-46	45	↑
ENSG00000136244	*IL6*	Interleukin 6	−2.46256	3.45793E-68	1.14219E-66	43	↓
ENSG00000100985	*MMP9*	Matrix metallopeptidase 9	2.49271	1.6074E-30	1.64089E-29	32	↑
ENSG00000174697	*LEP*	Leptin	−4.79532	3.2669E-200	4.5969E-197	30	↓
ENSG00000181092	*ADIPOQ*	Adiponectin, C1Q and collagen domain containing	−5.25949	4.1779E-160	1.278E-157	28	↓
ENSG00000175445	*LPL*	Lipoprotein lipase	−4.50595	7.0307E-190	6.3417E-187	25	↓
ENSG00000079435	*LIPE*	Lipase E, hormone sensitive type	−4.43709	2.6001E-167	1.0277E-164	24	↓
ENSG00000170323	*FABP4*	Fatty acid binding protein 4	−6.25417	1.5118E-173	7.3865E-171	24	↓
ENSG00000170345	*FOS*	Fos proto-oncogene, AP-1 transcription factor subunit	−3.1706	1.9545E-109	1.7018E-107	24	↓
ENSG00000118785	*SPP1*	Secreted phosphoprotein 1	2.093317	3.81647E-24	2.97026E-23	23	↑
ENSG00000108821	*COL1A1*	Collagen type I alpha 1 chain	2.302244	1.81421E-34	2.12028E-33	23	↑
ENSG00000120738	*EGR1*	Early growth response 1	−3.30574	7.4389E-137	1.1477E-134	21	↓
ENSG00000124253	*PCK1*	Phosphoenolpyruvate carboxykinase 1	−2.93033	1.4028E-135	2.0999E-133	20	↓
ENSG00000186847	*KRT14*	Keratin 14	−3.87354	3.78299E-50	7.39733E-49	20	↓
ENSG00000157404	*KIT*	KIT proto-oncogene, receptor tyrosine kinase	−3.18628	4.9103E-109	4.2441E-107	20	↓
ENSG00000062282	*DGAT2*	Diacylglycerol O-acyltransferase 2	−2.82748	4.32562E-84	2.10466E-82	19	↓
ENSG00000166819	*PLIN1*	Perilipin 1	−5.64203	5.604E-182	4.107E-179	18	↓
ENSG00000181856	*SLC2A4*	Solute carrier family 2-member 4	−2.43219	1.9755E-199	2.6729E-196	18	↓
ENSG00000171345	*KRT19*	Keratin 19	2.308081	2.30418E-36	2.86621E-35	18	↑

Information regarding the genes, including gene symbol, description, log-fold change, p-value, false discovery rate (FDR), degree of connectivity, and either upregulated or downregulated, is shown.

Table 2
Top 20 hub genes in M vs. N group from the PPI network.

ID	Gene symbol	Description	Log-fold change	P-value	FDR	Degree of connectivity	Up or down
ENSG00000136997	MYC	MYC proto-oncogene, bhlh transcription factor	−2.07121	0.00562	0.03931	59	↓
ENSG00000138160	KIF11	Kinesin family member 11	2.640848	1.74E-11	2.25E-09	52	↑
ENSG00000110092	CCND1	Cyclin D1	2.616773	8.53E-08	3.79E-06	52	↑
ENSG00000145386	CCNA2	Cyclin A2	2.72166	8.16E-12	1.14E-09	49	↑
ENSG00000164109	MAD2L1	Mitotic arrest deficient 2 like 1	2.403655	1.87E-09	1.34E-07	49	↑
ENSG00000091831	ESR1	Estrogen receptor 1	2.35876	0.002877	0.023246	48	↑
ENSG00000106462	EZH2	Enhancer of zeste 2 polycomb repressive complex 2 subunit	2.14774	2.81E-07	1.07E-05	47	↑
ENSG00000051180	RAD51	RAD51 recombinase	2.836993	2.3E-16	1.31E-13	46	↑
ENSG00000161800	RACGAP1	Rac GTPase activating protein 1	2.50692	8.92E-17	5.81E-14	46	↑
ENSG00000163808	KIF15	Kinesin family member 15	2.296105	6.93E-13	1.36E-10	45	↑
ENSG00000080986	NDC80	NDC80 kinetochore complex component	2.101671	2.07E-10	1.95E-08	44	↑
ENSG00000142731	PLK4	Polo like kinase 4	2.292129	5.46E-09	3.47E-07	44	↑
ENSG00000112742	TTK	TTK protein kinase	2.322704	1.69E-08	9.49E-07	42	↑
ENSG00000135476	ESPL1	Extra spindle pole bodies like 1, separase	2.223813	3.12E-11	3.7E-09	41	↑
ENSG00000125378	BMP4	Bone morphogenetic protein 4	−2.07565	3.57E-07	1.31E-05	40	↓
ENSG00000104147	OIP5	Opa interacting protein 5	2.62832	1.87E-15	8.44E-13	40	↑
ENSG00000111247	RAD51AP1	RAD51 associated protein 1	2.305299	1.19E-09	9.09E-08	40	↑
ENSG00000158402	CDC25C	Cell division cycle 25C	2.893235	8.54E-13	1.64E-10	40	↑
ENSG00000065328	MCM10	Minichromosome maintenance 10 replication initiation factor	2.176317	1.78E-13	4.26E-11	40	↑
ENSG00000176890	TYMS	Thymidylate synthetase	2.348964	1.23E-06	3.78E-05	39	↑

Information regarding the genes, including gene symbol, description, log-fold change, p-value, false discovery rate (FDR), degree of connectivity, and either upregulated or downregulated, is shown.

genes (Fig. 5d–f). Supplementary Figure S1 presents the overlap between the top 20 hub genes and each cluster model.

3.3. Functional enrichment analysis identified pathways associated with BC metastasis

Figs. 6–8 illustrate the top significant terms from the GO functional annotation and KEGG pathways of the top twenty hub genes in both P vs. N and M vs. N analyses. The top 10 GO terms and KEGG pathways for the PPI networks of P vs. N and M vs. N groups are shown in Supplementary Tables S8–S11 and Figures S2–S6.

3.3.1. GO functional annotation of the top twenty hub genes in the P vs. N PPI group

MF ontology analysis (Fig. 6a–c, Tables S9) revealed that 71.43% of the terms had identical protein-binding groups. Furthermore,

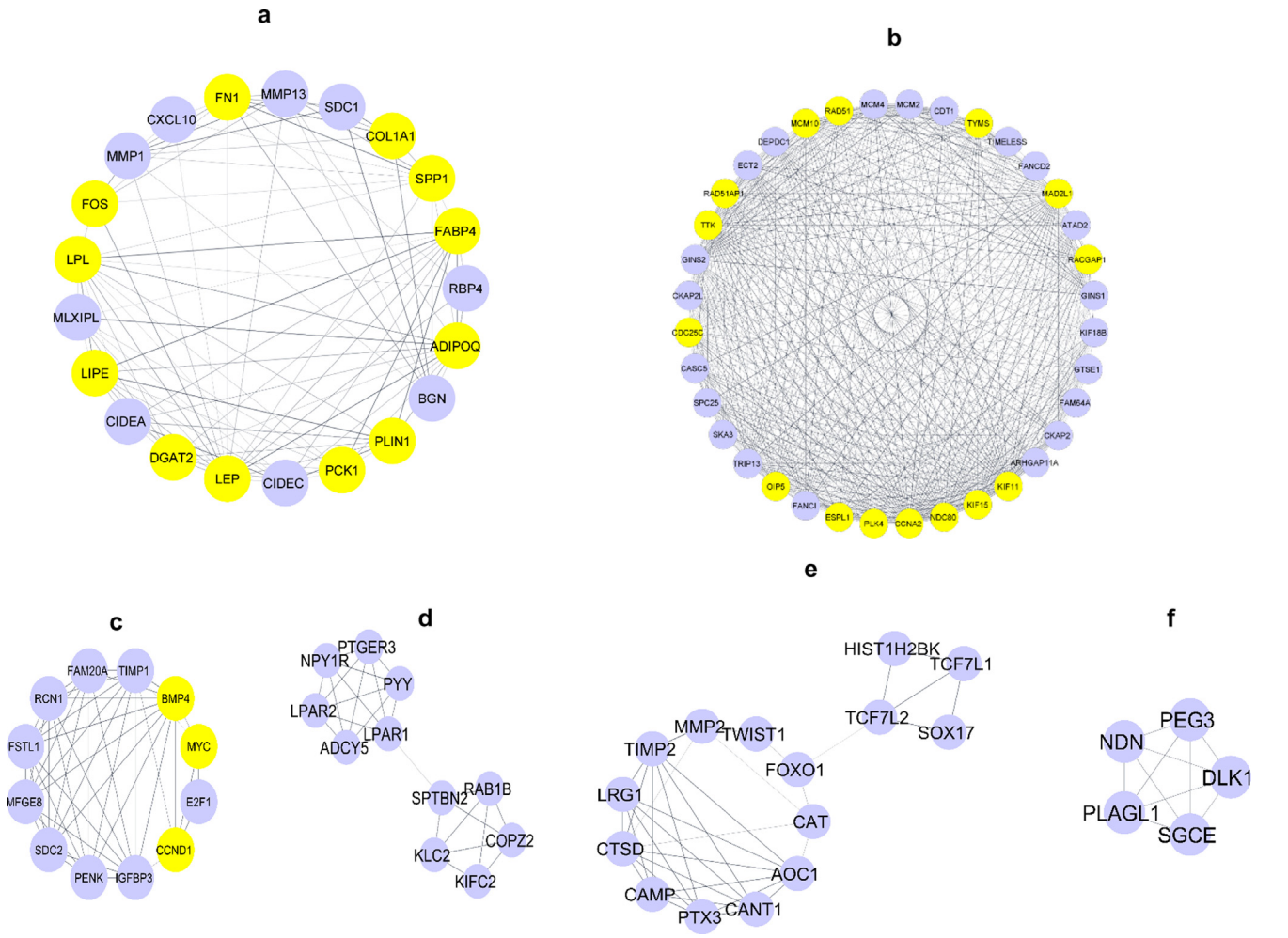

Fig. 5. Molecular complex detection (MCODE) clusters for selected significant modules from protein-protein interaction (PPI) networks in both P vs. N and M vs. N groups. (**a**) cluster 1 in P vs. N group. (**b**–**f**) The five significant modules from PPI networks of M vs. N group; (b) cluster 1 in M vs. N group, (c) cluster 2 in M vs. N group, (d) cluster 3 in M vs. N group, (e) cluster 4 in M vs. N group, and (f) cluster 5 in M vs. N group. Yellow highlights the genes of the seeds that were common with the top 20 hub genes in each group; the MCODE cluster is based on a cut-off criterion ≥ 5. Figures were generated post analysis using MCODE tool and search tool for the retrieval of interacting genes (STRING) in Cytoscape. P vs. N, primary tumour tissue samples obtained from patients with infiltrating duct carcinoma compared with normal breast tissue samples; M vs. N, metastatic tumour tissue samples obtained from axillary lymph nodes of patients with primary infiltrating duct carcinoma of the breast compared with normal breast tissue samples.

14.29% of the identified terms were in the protein-containing complex binding group and 14.29% of the terms in the protein-binding group. In CC ontology analysis (Fig. 6d–f, Tables S9), 66.67% of the identified terms were in the extracellular space group. In addition, 22.22% of the identified terms were in the endoplasmic reticulum group and 11.11% of the terms in the cytoplasm group. In BP ontology analysis (Fig. 6g–I, Tables S9), 54.55% of the terms were found in the triglyceride metabolic process group and 45.45% of the terms in the positive regulation of chemokine production group.

3.3.2. GO functional annotation of the top 20 hub genes in the M vs. N PPI group

In the MF ontology analysis (Fig. 7a–c, Tables S10), 28.57% of the terms were found in the anion binding group. In total, 28.57% of the terms were associated with the activity of a transferase responsible for the transfer of phosphorus-containing compounds. In the binding group, 14.29% of the terms were identical. Furthermore, 14.29% of the terms were found in the enzyme-binding group and 14.29% of the terms were in the protein-binding group. In the CC ontology analysis (Fig. 7d–f, Tables S10), 28.57% of the terms were determined to be enriched in the intracellular organelle lumen group. Moreover, 28.57% of the terms were identified in the chromosome group, 14.29% of the terms were identified in the intracellular organelle group, and 14.29% were identified in the microtubule cytoskeleton group. An additional 14.29% of the terms were identified in the cytosol group.BP ontology analysis (Fig. 7g–i, Tables S10) identified 53.12% of the terms enriched in mitotic cell cycle phase transition. In addition, 28.12% of the genes were associated with negative regulation of the cell cycle. Moreover, 15.62% of the terms were enriched in regulation of cell cycle. The remaining 3.12% of the terms were associated with cell cycle.

3.3.3. Pathway enrichment of the top twenty hub genes in the P vs. N PPI group

The top 10 KEGG pathways for the top twenty hub genes in the P vs. N PPI group (Fig. 8a) were the PPAR signalling pathway (five genes), AMPK signalling pathway (five genes), adipocytokine signalling pathway (four genes), PI3K-AKT signalling pathway (six genes), regulation of lipolysis in adipocytes (three genes), pathways in cancer (five genes), CM-receptor interaction (three genes), amoebiasis (three genes), Toll-like receptor signalling pathway (three genes), and TNF signalling pathway (three genes).

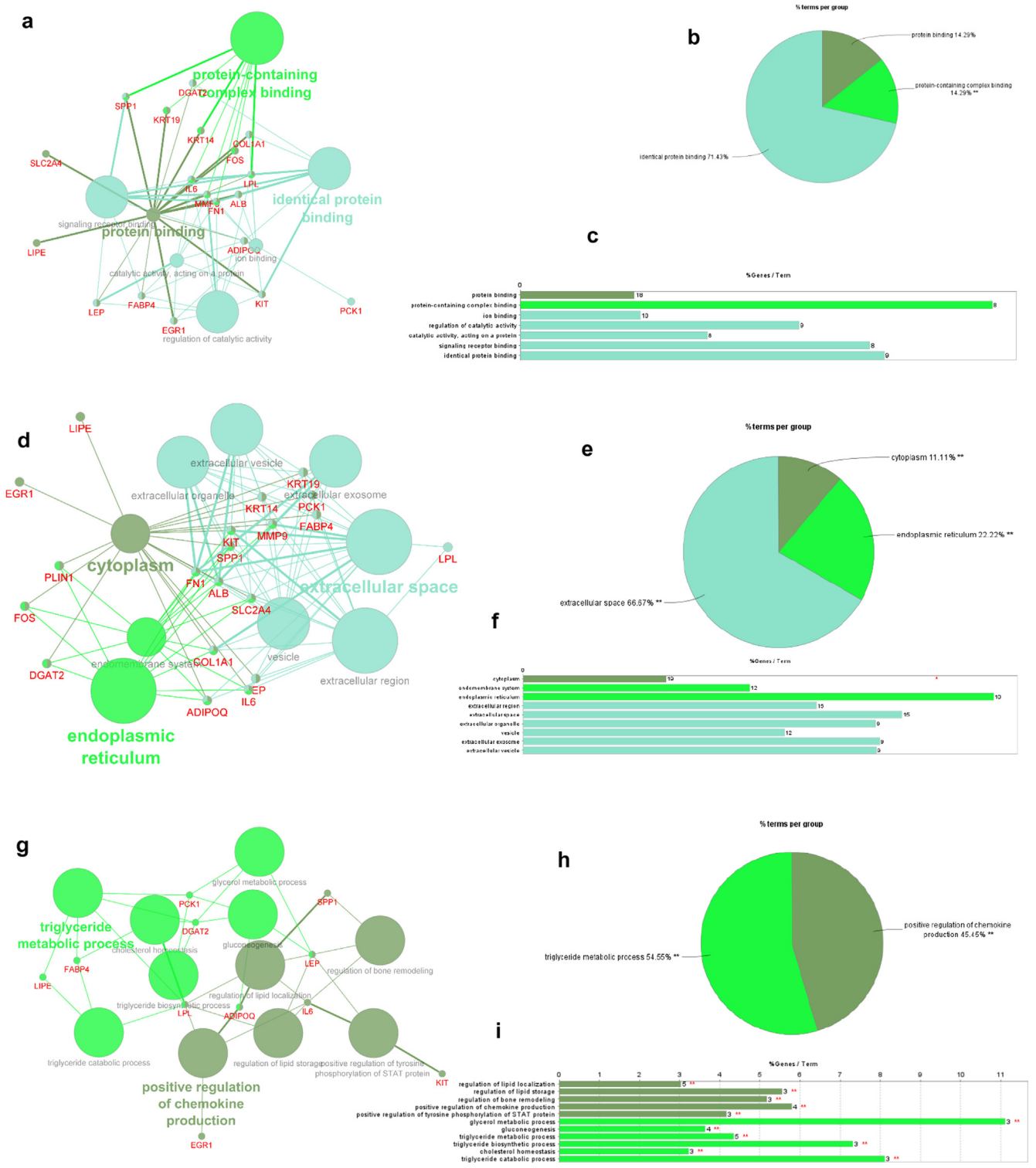

Fig. 6. Gene ontology (GO) term analysis of the top 20 hub genes in P vs. N group. (**a**–**c**) Molecular function category (MF). (**d**–**f**) Cellular component category (CC). (**g**–**i**) Biological processes category (BP). (a, d, g) Networks for GO terms that were unique to the top 20 hub genes in P vs. N group for each GO category. (b, e, h) Pie chart with % terms per group, including particular terms for the top 20 hub genes in each GO category from P vs. N group. (c, f, i) Bars indicate the % of genes per term associated with each GO category from P vs. N group. Figures were generated using the ClueGO and CluePedia plugins in Cytoscape. P vs. N, primary tumour tissue samples obtained from patients with infiltrating duct carcinoma compared with normal breast tissue samples; M vs. N, metastatic tumour tissue samples obtained from axillary lymph nodes of patients with primary infiltrating duct carcinoma of the breast compared with normal breast tissue samples.

3.3.4. Pathway enrichment of the top 20 hub genes in the P vs. N PPI group

The top KEGG pathways for the top 20 hub genes in the M vs. N PPI group (Fig. 8b) included cell cycle (seven genes), thyroid hormone signalling pathway (four genes), progesterone-mediated oocyte maturation (three genes), microRNAs in cancer (three genes), oocyte meiosis (three genes), hepatitis B (three genes), Hippo signalling pathway (three genes), pathways in cancer (four genes), proteoglycans in cancer (three genes), thyroid cancer (two genes), bladder cancer (two genes), HTLV-I infection (three

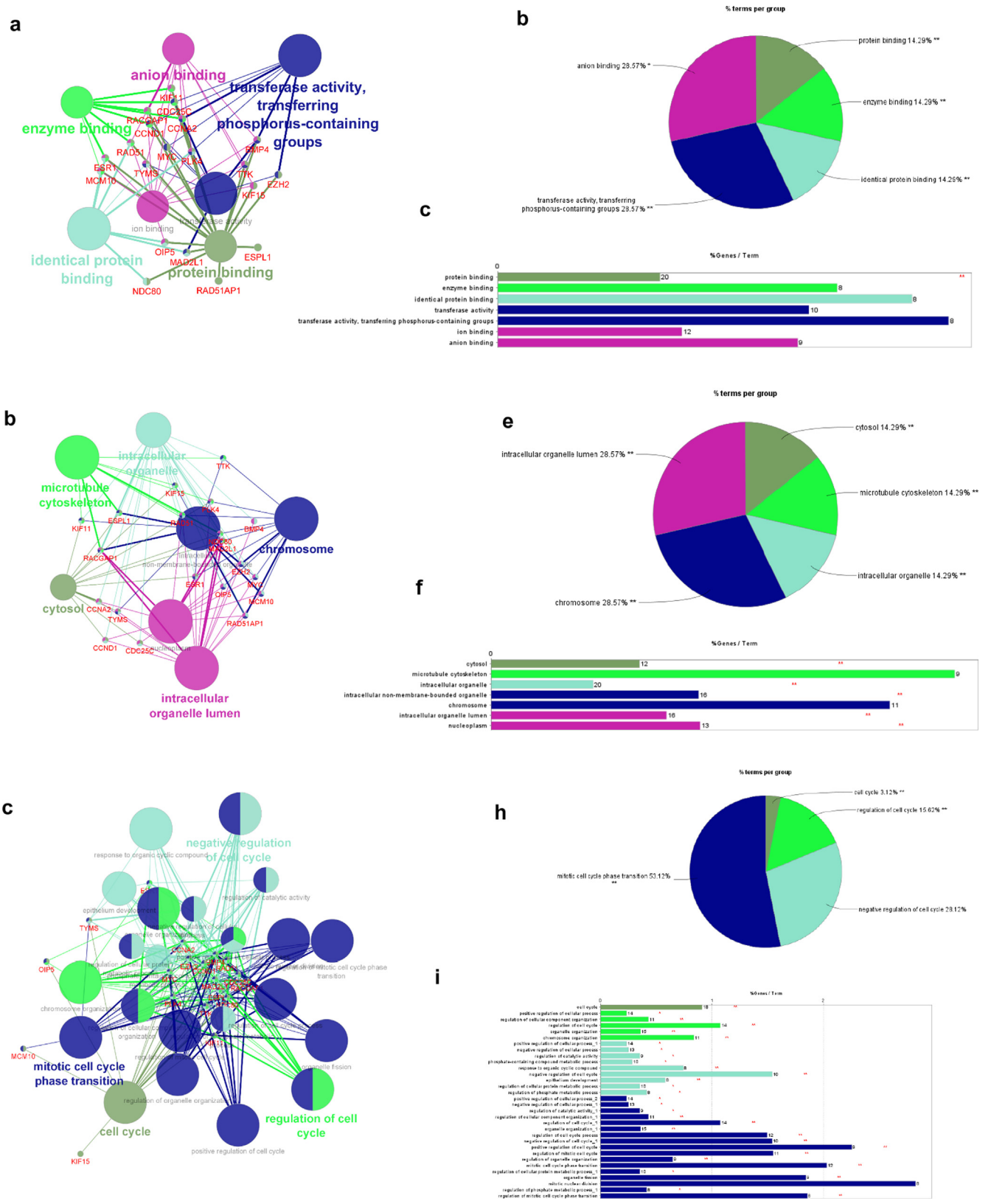

Fig. 7. Gene ontology (GO) term analysis of the top 20 hub genes in M vs. N group. (**a**–**c**) Molecular function category (MF). (**d**–**f**) Cellular component category (CC). (**g**–**i**) Biological processes category (BP). (a, d, g) Networks for GO terms that were unique to the top 20 hub genes in M vs. N group for each GO category. (b, e, h) Pie chart with % terms per group, including particular terms for the top 20 hub genes in each GO category from M vs. N group. (c, f, i) Bars indicate the % of genes per term associated with each GO category from M vs. N group. Figures were generated using the ClueGO and CluePedia plugins in Cytoscape. P vs. N, primary tumour tissue samples obtained from patients with infiltrating duct carcinoma compared with normal breast tissue samples; M vs. N, metastatic tumour tissue samples obtained from axillary lymph nodes of patients with primary infiltrating duct carcinoma of the breast compared with normal breast tissue samples.

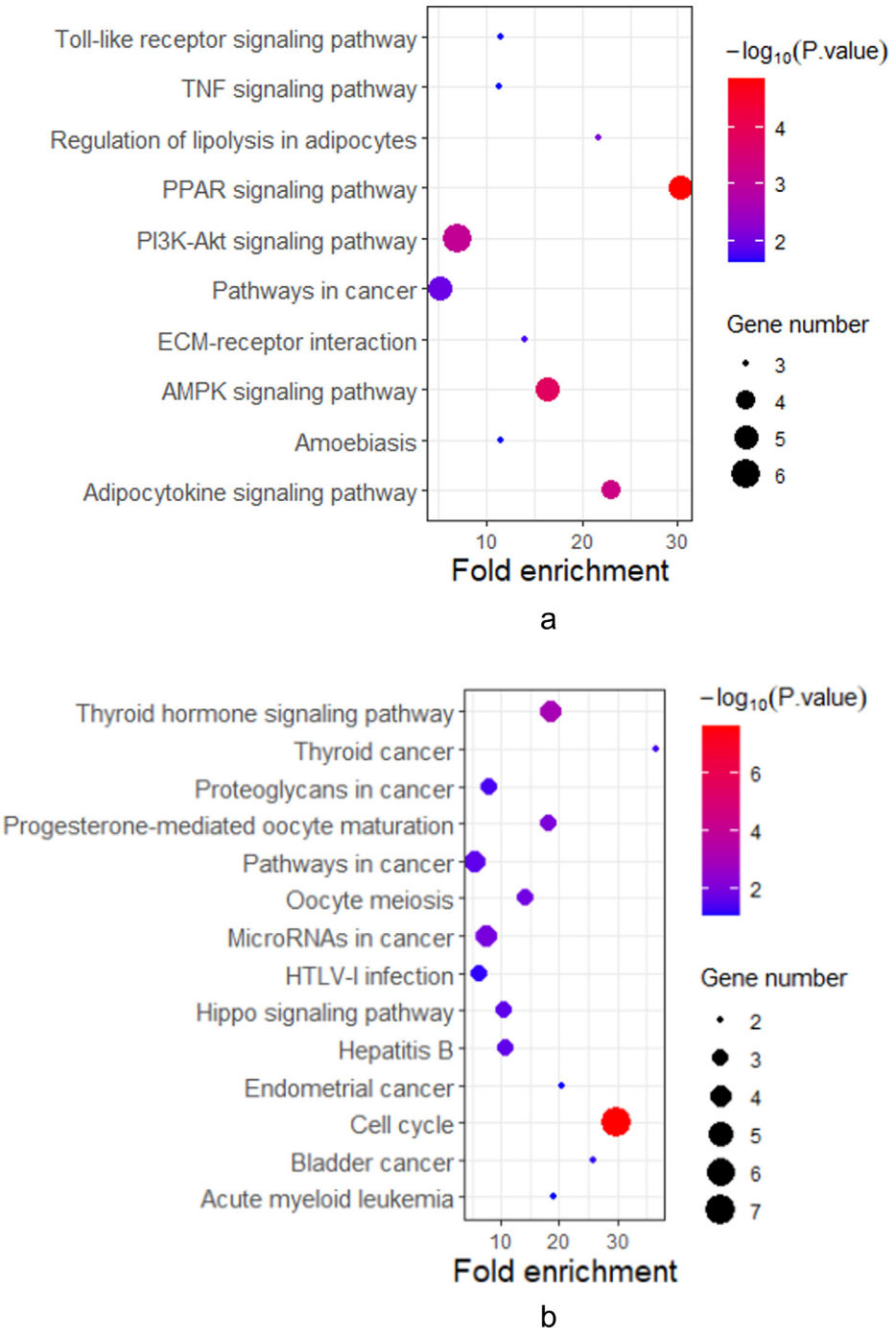

Fig. 8. Bar graphs of top 10 pathways identified in Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis that were significantly enriched for the top 20 hub genes. Each point corresponds to a KEGG pathway, y-axis displays the pathway's name, x-axis displays the fold enrichment, dot size corresponds to the number of genes, the colour bar displays the –log10 (P-value), red indicates a high value, and blue indicates a low value. (**a**) P vs. N, (**b**) M vs. N. The greater the fold enrichment, the greater the significance of DEG enrichment in this pathway. P vs. N, primary tumour tissue samples obtained from patients with infiltrating duct carcinoma compared with normal breast tissue samples; M vs. N, metastatic tumour tissue samples obtained from axillary lymph nodes of patients with primary infiltrating duct carcinoma of the breast compared with normal breast tissue samples.

genes), endometrial cancer (two genes), and acute myeloid leukaemia (two genes).

3.4. Correlation analysis revealed candidate hub genes associated with overall survival in BC

Next, the correlations among the expression of the top twenty candidate hub genes and the OS of patients with BC in both P vs. M and M vs. N groups were evaluated using Kaplan–Meier curves to ascertain the hub genes' prognostic significance (Figs. 9 and 10).

3.4.1. Candidate hub genes associated with OS in P vs. N group

The GEPIA results (Fig. 9) indicated that the upregulated expression of *FN1*, *SPP1*, *COL1A1*, and *KRT19* and the downregulated expression of *ALB*, *IL6*, *FOS*, *EGR1*, *PCK1*, *KRT14*, *KIT*, and *DGAT2* markers were correlated with poor OS, which is in accordance with

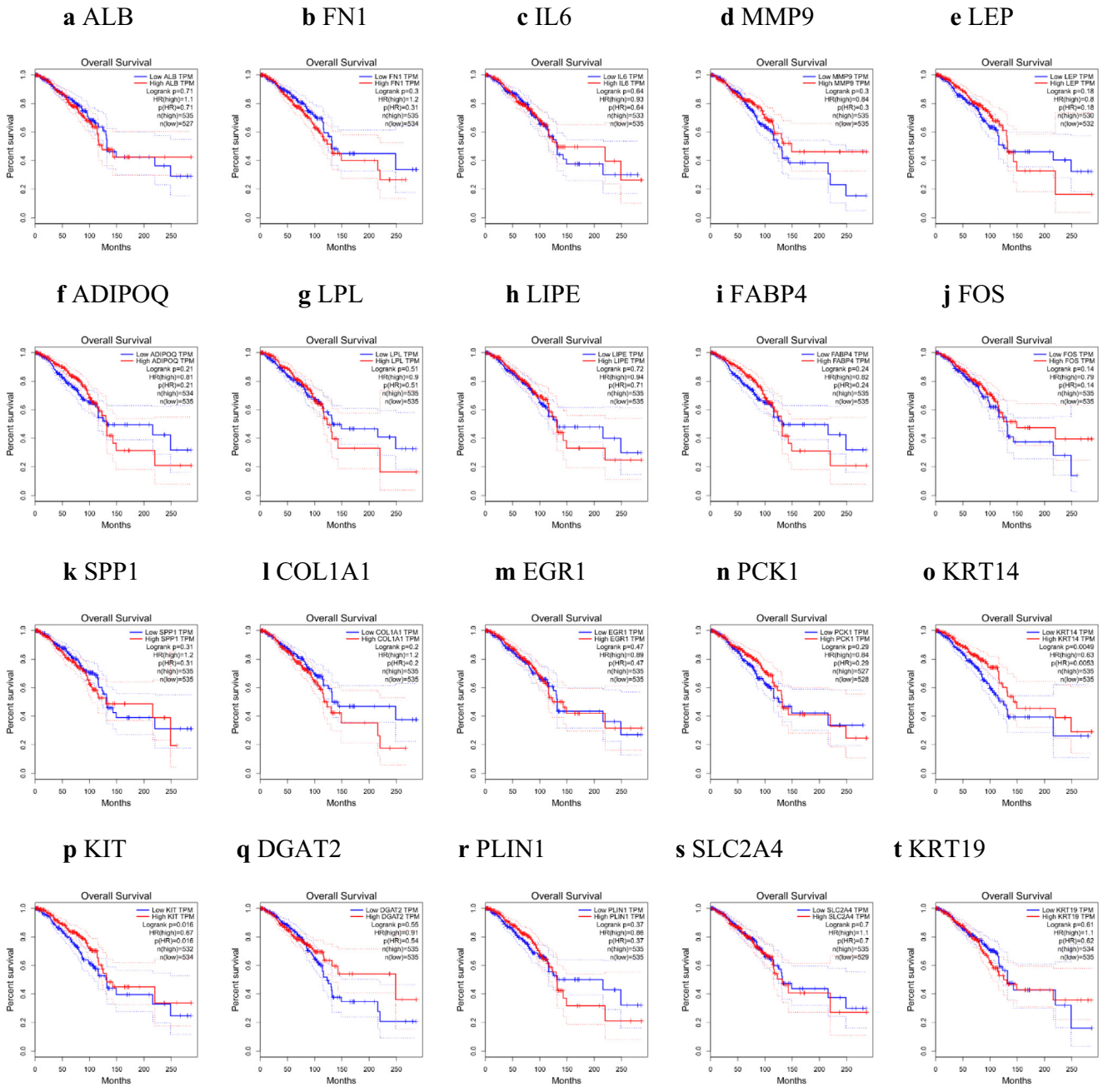

Fig. 9. Overall survival (OS) analysis of the 20 hub gene candidates identified in P vs. N group using the Kaplan–Meier curve and gene expression profiling interactive analysis (GEPIA) to correlate gene expression to time and percentage of survival. Low expression group represented in blue, high expression group represented in red, and number of patients represented as n. A difference of $p < 0.05$ was considered significant, and median was used as the group cut-off; the high and low cut-offs were 50%. Hazard ratio (HR) was calculated based on the Cox PH model, 95% confidence interval (CI) was added to the plot as a dotted line, and BRCA was used as dataset selection. (**a**) *ALB*, (**b**) *FN1*, (**c**) *IL6*, (**d**) *MMP9*, (**e**) *LEP*, (**f**) *ADIPOQ*, (**g**) *LPL*, (**h**) *LIPE*, (**i**) *FABP4*, (**j**) *FOS*, (**k**) *SPP1*, (**l**) *COL1A1*, (**m**) *EGR1*, (**n**) *PCK1*, (**o**) *KRT14*, (**p**) *KIT*, (**q**) *DGAT2*, (**r**) *PLIN1*, (**s**) *SLC2A4*, and (**t**) *KRT19*. P vs. N, primary tumour tissue samples obtained from patients with infiltrating duct carcinoma compared with normal breast tissue samples.

the log-fold change results generated using the mRNASeq platform in this study. Unexpectedly, the GEPIA results revealed higher expression of *LEP*, *ADIPOQ*, *LPL*, *LIPE*, *FABP4*, *PLIN1*, and *SLC2A4* and lower expression of *MMP9* than the mRNASeq results, which might correlate with poor OS. The criteria for identifying prognostic risk factors included: log-rank $p < 0.05$ and an HR score > 1 in upregulated genes and < 1 in downregulated genes (Xu et al., 2020). Based on log-rank p and HR scores, only two genes from the candidate hub genes in the P vs. N group were considered prognostic risk factors; these were *KRT14* (HR = 0.63, log-rank p = 0.0049) and *KIT* (HR = 0.67, log-rank p = 0.016), and the expression of both was downregulated.

3.4.2. No significant candidate hub genes were associated with OS in M vs. N group

The GEPIA findings (Fig. 10) indicated that the upregulated expression of *CCND1*, *CCNA2*, *MAD2L1*, *ESR1*, *RAD51*, *TTK*, *ESPL1*, *RAD51AP1*, *CDC25C*, *MCM10*, and *TYMS* expression and the downregulated expression of *MYC* might be correlated with poor OS; these results were supported by mRNA-Seq-generated log-

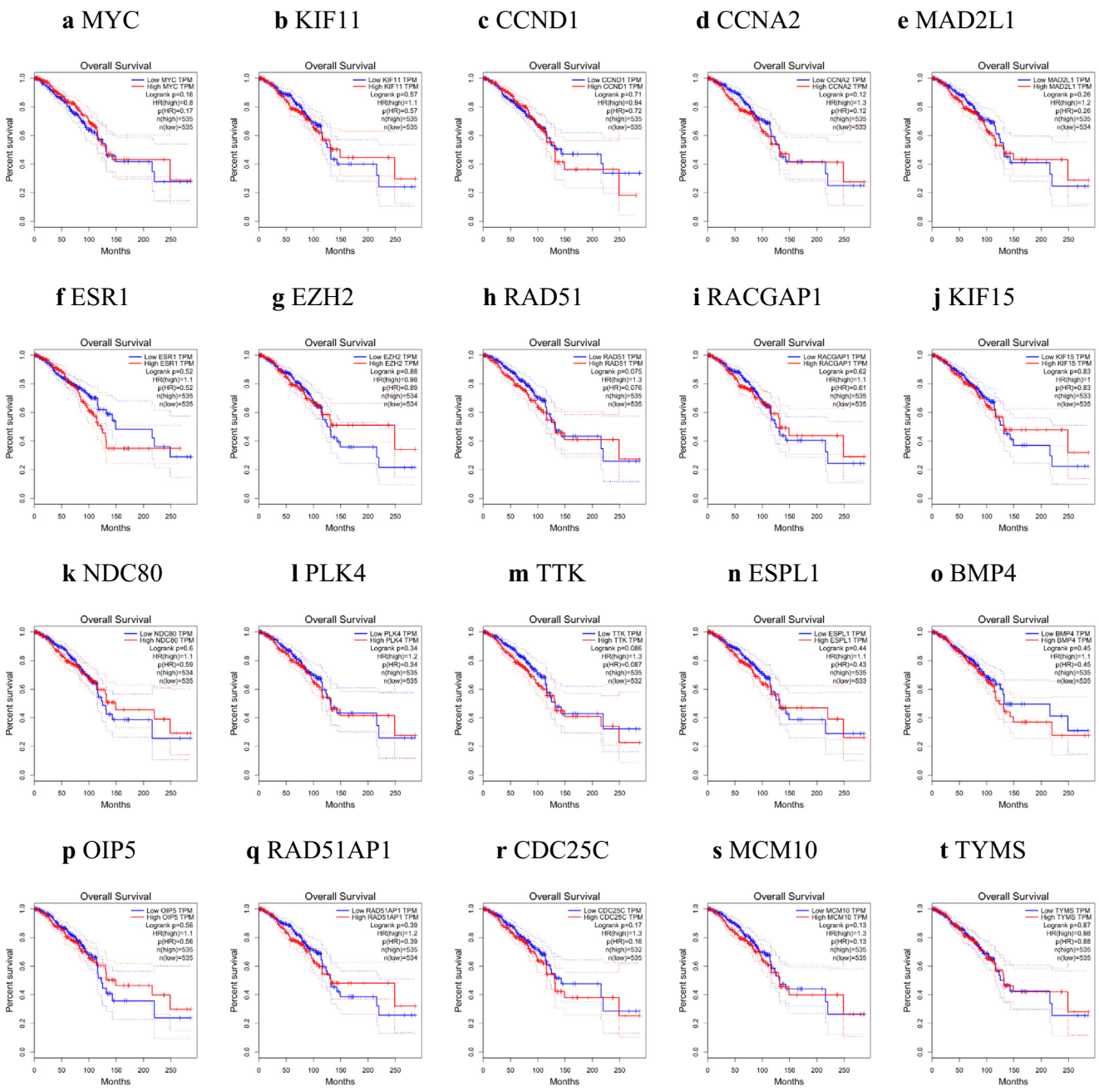

Fig. 10. Overall survival (OS) analysis on the 20 hub gene candidates identified in M vs. N group using the Kaplan–Meier curve and gene expression profiling interactive analysis (GEPIA) to correlate gene expression to time and percentage of survival. Low expression group represented in blue, high expression group represented in red, and number of patients represented as n. A difference of $p < 0.05$ was considered significant, and median was used as the group cut-off; the high and low cut-offs were 50%. Hazard ratio (HR) was calculated based on the Cox PH model, 95% confidence interval (CI) was added to the plot as a dotted line, and BRCA was used as dataset selection. (**a**) *MYC*, (**b**) *KIF11*, (**c**) *CCND1*, (**d**) *CCNA2*, (**e**) *MAD2L1*, (**f**) *ESR1*, (**g**) *EZH2*, (**h**) *RAD51*, (**i**) *RACGAP1*, (**j**) *KIF15*, (**k**) *NDC80*, (**l**) *PLK4*, (**m**) *TTK*, (**n**) *ESPL1*, (**o**) *BMP4*, (**p**) *OIP5*, (**q**) *RAD51AP1*, (**r**) *CDC25C*, (**s**) *MCM10*, and (**t**) *TYMS*. M vs. N, metastatic tumour tissue samples obtained from axillary lymph nodes of patients with primary infiltrating duct carcinoma of the breast compared with normal breast tissue samples.

transformed results. Unexpectedly, GEPIA results showed higher *BMP4* expression and lower *KIF11*, *EZH2*, *RACGAP1*, *KIF15*, *NDC80*, *PLK4*, *BMP4*, and *OIP5* expression than the mRNA-Seq results, which correlated with poor OS. Based on the log-rank p and HR scores, these nine candidate hub genes in the M vs. N group could be considered as prognostic risk factors. However, none of these genes were statistically significant, with only two of them approaching significance: *RAD51* (HR = 1.3, log-rank p = 0.075) and *TTK* (HR = 1.3, log-rank p = 0.087); the expression of both was upregulated. Thus, no prognostic risk factors were clearly identified in this study.

3.5. Validation of mRNA and protein expression of the identified potential prognostic risk factors

The Human Protein Atlas database was utilized to examine the protein expression levels of the four hub genes that were considered as potentially prognostic risk factors based on OS. Notably, *KRT14* (Fig. 11a) and *KIT* (Fig. 11b) were not expressed in BC tissues, whereas moderate *KRT14* levels and high *KIT* levels were observed in normal breast tissues, implying that hub gene expression was downregulated at the transcriptional and translational levels in patients with BC. Additionally, *RAD51* (Fig. 11c) and *TTK*

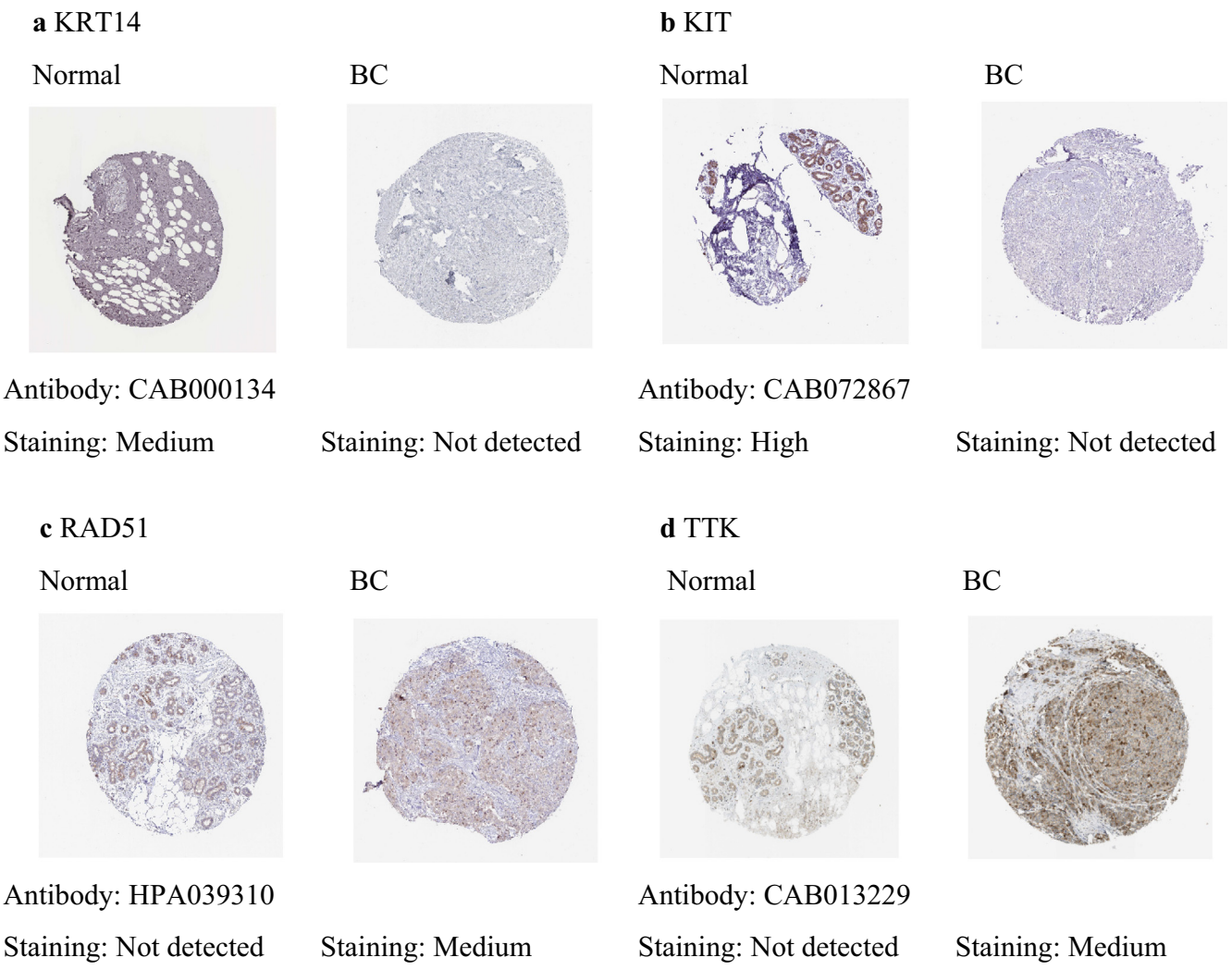

Fig. 11. Immunohistochemistry staining images. (**a**) KRT14, (**b**) RAD51, (**c**) IL-6, and (**d**) TTK expression in breast cancer (BC) tissues versus that in normal breast tissues from The Human Protein Atlas database (immunohistochemistry).

(Fig. 11d) were not detected in normal breast tissues but were detected at moderate levels in BC tissues, indicating that the transcript and protein levels of both hub genes were increased in patients with BC.

4. Discussion

Patients with BC who have axillary lymph node metastases have a greater chance of early relapse and a shorter OS compared to those with BC with primary non-metastatic tumours (Panel, 2001). The mechanisms underlying distant metastases are complex and involve a diverse array of genes and signalling pathways (Pachmayr et al., 2017, Nathanson, 2003). Hence, the identification of predictive biomarkers that determine the extent of cancer spread and subsequent development of novel therapeutics could be highly effective in eradicating metastatic cancer. Determination of such biomarkers that can identify the transition point from the initial stage to metastasis would require high-throughput genome-wide association analysis. Numerous research have attempted to explain the molecular pathways behind the transformation from primary cancer to metastasis by comparing gene expression profiles of primary and healthy tissues as well as of primary and metastatic tissues; however, these studies focused exclusively on the changes in the tumour microenvironment and differences in gene expression and proteins between these groups (Ikemura et al., 2017, Hao et al., 2004, Mimori et al., 2005, Feng et al., 2007, Vecchi et al., 2008, Ellsworth et al., 2009).

This study revealed many DEGs common to both primary and metastatic tumours as well as several DEGs that were characteristic of each type. Additionally, the top 20 candidate hub genes uniquely expressed in tissues of primary and lymph node metastatic BC were identified to determine the molecular changes between them. Among these top 20 candidate hub genes, *KRT14* and *KIT* were identified in primary BC tissues in the P vs. N group, whereas *RAD51* and *TTK* were detected in lymph node metastasis BC tissues in the M vs. N group. These four genes were considered to be prognostic risk factors based on the significant association between their expression and OS. The transcriptional and translational levels of *KRT14* and *KIT* were higher, whereas those of *RAD51* and *TTK* were lower, in patients with BC.

KRT14 is a filamentous intermediate protein that is predominantly produced by epithelial progenitor cells (Chu et al., 2001). This cell population exhibits migratory behaviour, replication, and differentiation, and controls branching morphogenesis throughout organ development (Papafotiou et al., 2016, Abashev et al., 2017, Rock et al., 2009). The increased expression of KRT14 in primary BC tissues (P vs. N) is comparable to that observed in

a previous study, wherein a > 100-fold higher expression of *KRT14* was found in RNA isolated from 20 patients with primary BC (Ellsworth et al., 2009). Similarly, Hanley *et al.* found a significant correlation between an extracellular matrix organisation module mediated by stromal cells and *KRT14* expression. *KRT14* expression and the basal transcription factor *TP63* are upregulated in cancer cells with genetic autophagy inhibition. KRT14–TP63 expression has been found to be highly pro-metastatic in human and mouse mammary tumours (Marsh and Debnath, 2020). Similarly, an *in vitro* model of epithelial ovarian cancer metastasis and imaging mass spectrometry revealed *KRT14* expression to be key for identifying the invasive potential of ovarian cancer cells (Bilandzic et al., 2019). Furthermore, although *KRT14* is required for invasion, it has no effect on the viability or proliferation of cells (Bilandzic et al., 2019). *KRT14* is likely to perform a distinctive role during invasion, additionally to maintaining cytoskeletal stability and integrity (Montor et al., 2018, Karimnia et al., 2021, Pearson, 2019, Werner et al., 2020). *KRT14* is expressed in numerous types of tumours, particularly in invasive tumour cells as opposed to non-invasive tumour cells (Chu et al., 2001, Papafotiou et al., 2016, Cheah et al., 2015, Volkmer et al., 2012, Cheung et al., 2016, Cheung et al., 2013, Lichtner et al., 1991, Gordon et al., 2003, Petrocca et al., 2013). In fact, in 2020, a study demonstrated that the expression of a 34-gene profile, including *KRT14*, represents a favourable prognostic factor for squamous non-small cell lung carcinoma (Theelen et al., 2020). In another study, a comprehensive bioinformatics analysis was conducted to identify potential genes associated with prognosis and treatment in BC in four GEO datasets. They identified seven genes, including *KRT14*, as critical prognostic candidate genes that were strongly related with OS in BC (Xu et al., 2020).

KIT, a tyrosine kinase of transmembrane receptors, plays a vital function in cell proliferation, survival, and migration (Liu et al., 2017, Hernández-Rojas et al., 2022, Wintheiser and Silberstein, 2021, Montor et al., 2018). Notably, tumour progression is linked to the downregulation of *c-KIT* expression, which is thought to be involved in human BC development during the initial stages (Janostiak et al., 2018). Talaiezadeh *et al.* evaluated *KIT* expression in 60 BC and normal breast specimens via immunohistochemistry and concluded that there is an association between downregulation of *KIT* and the transformation of breast epithelial cells into cancer cells. KIT has also been assessed for its prognostic and predictive utility in patients with high-risk BC. Specifically, its expression was evaluated in 236 patients using tissue microarrays and its downregulation was detected in 12% of patients with BC. Furthermore, its expression was determined to be strongly related with poorer OS (Diallo et al., 2006). Thus, KIT expression may be a separate negative prognostic factor in patients with high-risk recurrence of BC. Comparable to the findings of the current study, a previous study of approximately 667 patients with BC showed that individuals with basal-like BC (28%) and those with nodal metastases (218%) had higher KIT positive tumours than healthy controls. They concluded that KIT may be a prognostic marker for individuals with basal-like BC and a potential biological target for treatment (Kashiwagi et al., 2013).

RAD51 is necessary for the homologous recombination mechanism to repair damaged DNA (Grundy et al., 2020, Inano et al., 2017, Gachechiladze et al., 2017). Thus, *RAD51* mutations are directly correlated with genome instability and a predisposition for cancer development (Prakash et al., 2015, Sullivan and Bernstein, 2018). Moreover, RAD51 has been found to have novel functions in pancreatic cancer that, mechanistically, may involve the regulation of aerobic glycolysis, which is associated with OS outcome (Zhang et al., 2019b). In fact, patients with pancreatic cancer and elevated RAD51 levels show poorer prognosis, suggesting that RAD51 serves as a prognostic biomarker for pancreatic cancer and regulates cell proliferation (Zhang et al., 2019b). Nonetheless, *RAD51* expression in cancer continues to be investigated both clinically and biologically. Specifically, RAD51 has come under intense scrutiny owing to its involvement in tumour progression and resistance to chemotherapy (Raderschall et al., 2002). Moreover, *RAD51* is overexpressed in invasive ductal BC, and the degree of overexpression is connected to the histological categorisation of BC (Maacke et al., 2000). Additionally, another study indicated that RAD51 expression may be prognostic for individuals with surgically treated non-small cell lung cancer, as the lower survival of patients with non-small cell lung cancer with high levels of RAD51 expression was associated with an increased proclivity of tumour cells for survival, therapeutic resistance, and resistance to apoptosis. These findings suggest that RAD51 expression at a high level could be used as a prognostic marker of survival in patients with BC.

TTK is an important member of a number of mitotic stages in a wide range of eukaryotic cells (Wang et al., 2018, Pangou and Sumara, 2021). It is a major kinase implicated in the localisation of kinetochores and the spindle assembly checkpoint (Liu et al., 2020, Pachis and Kops, 2018). *TTK* is overexpressed in several different kinds of cancer and protects against chromosomal anomalies (Xie et al., 2017, Maire et al., 2013, Liu et al., 2015, Suyal et al., 2022, Yao et al., 2022, Jiao et al., 2022). According to earlier studies, TTK is upregulated in BC tissue and cells, notably in HER2-positive and triple-negative BC subtypes (Daniel et al., 2011, Salvatore et al., 2007, Tannous et al., 2013, Cui and Guadagno, 2008). Previously published research addressed the role of TTK in cancer utilizing *in vitro* pharmacological inhibitors. For instance, in the colorectal and glioblastoma *in vitro* models, the inhibition of TTK expression pharmacologically decreases cell viability, resulting in abnormal cell cycle progression, elevated aneuploidy, and induced apoptosis (Yao et al., 2021). TTK activity is crucial for pancreatic cancer cell proliferation, making TTK a potential therapeutic target for new treatments (Liu et al., 2021). Additionally, TTK was evaluated as a potential prognostic biomarker for triple-negative BC and it was demonstrated that increased TTK expression is associated with improved disease-free survival and OS, with a better prognosis (Xu et al., 2016). Similarly, the decreased *TTK* mRNA expression was positively correlated with worse OS, elevated incidence of metastasis, and shorter disease-free survival, all of which are indicators of poor prognosis (Maire et al., 2013). Another study using RNA microarray showed that more aggressive tumours have higher *TTK* expression, resulting in extremely low survival of less than two years (Al-Ejeh et al., 2014). TTK is variably expressed in normal and cancerous tissues, highlighting its viability as a genetic marker for diagnosis. Nevertheless, the fact that it correlates strongly with relapse and overall survival implies that it also might function as an independently prognostic marker (Xie et al., 2017). Based on the findings of this study, it is recommended to use TTK expression as a predictive marker in patients with IDC to aid in diagnosis as well as for the creation of individualised treatment plans for patients.

5. Conclusion

The current study used a novel approach to identify DEGs associated with the prognosis of primary and metastatic BC. Of the top 20 hub genes identified, four genes [*KRT14*, *KIT* (upregulated in BC) and *RAD51*, *TTK* (downregulated in BC)] could be considered for risk prognosis based on OS. However, the clinical importance of *KRT14*, *KIT*, *RAD51*, and *TTK* as prognostic factors should be further investigated prospectively. Future analysis of these four proposed prognostic genes may also provide insights into the molecular differences between primary breast tumours and metastatic lymph

node metastases. In addition, the genes whose expression is altered in BC metastases may be used to design innovative therapies that precisely target metastatic colonisation.

Funding

This work was supported by the Taif University Researchers Supporting Project, Taif University, Taif, Saudi Arabia (Grant number TURSP-2020/202).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sjbs.2022.103318.

References

Abashev, T.M., Metzler, M.A., Wright, D.M., Sandell, L.L., 2017. Retinoic acid signaling regulates Krt5 and Krt14 independently of stem cell markers in submandibular salivary gland epithelium. Dev. Dyn. 246, 135–147.

AL-EJEH, F., SIMPSON, P., SANUS, J., KLEIN, K., KALIMUTHO, M., SHI, W., MIRANDA, M., KUTASOVIC, J., RAGHAVENDRA, A. & MADORE, J. 2014. Meta-analysis of the global gene expression profile of triple-negative breast cancer identifies genes for the prognostication and treatment of aggressive breast cancer. *Oncogenesis,* 3**,** e100-e100.

Albogami, S.M., Asiri, Y., Asiri, A., Alnefaie, A.A., Alnefaie, S., 2021. Effects of neoadjuvant therapies on genetic regulation of targeted pathways in ER+ primary ductal breast carcinoma: a meta-analysis of microarray datasets. Saudi Pharma. J. 29, 656–669.

Andreopoulou, E., Hortobagyi, G.N., 2008. Prognostic factors in metastatic breast cancer: successes and challenges toward individualized therapy. J. Clin. Oncol. 26, 3660–3662.

Bilandzic, M., Rainczuk, A., Green, E., Fairweather, N., Jobling, T.W., Plebanski, M., Stephens, A.N., 2019. Keratin-14 (KRT14) positive leader cells mediate mesothelial clearance and invasion by ovarian cancer cells. Cancers 11, 1228.

Bindea, G., Galon, J., Mlecnik, B., 2013. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. Bioinformatics (Oxford, England) 29, 661–663.

Cheah, M.T., Chen, J.Y., Sahoo, D., Contreras-Trujillo, H., Volkmer, A.K., Scheeren, F.A., Volkmer, J.-P., Weissman, I.L., 2015. CD14-expressing cancer cells establish the inflammatory and proliferative tumor microenvironment in bladder cancer. Proc. Natl. Acad. Sci. 112, 4725–4730.

Cheung, K.J., Gabrielson, E., Werb, Z., Ewald, A.J., 2013. Collective invasion in breast cancer requires a conserved basal epithelial program. Cell 155, 1639–1651.

Cheung, K.J., Padmanaban, V., Silvestri, V., Schipper, K., Cohen, J.D., Fairchild, A.N., Gorin, M.A., Verdone, J.E., Pienta, K.J., Bader, J.S., 2016. Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. Proc. Natl. Acad. Sci. 113, E854–E863.

Chu, P., Lyda, M., Weiss, L., 2001. Cytokeratin 14 expression in epithelial neoplasms: a survey of 435 cases with emphasis on its value in differentiating squamous cell carcinomas from other epithelial tumours. Histopathology 39, 9–16.

Cianfrocca, M., Goldstein, L.J., 2004. Prognostic and predictive factors in early-stage breast cancer. Oncologist 9, 606–616.

Cilibrasi, C., Papanastasopoulos, P., Samuels, M., Giamas, G., 2021. Reconstituting immune surveillance in breast cancer: molecular pathophysiology and current immunotherapy strategies. Int. J. Mol. Sci. 22, 12015.

Ciriello, G., Gatza, M.L., Beck, A.H., Wilkerson, M.D., Rhie, S.K., Pastore, A., Zhang, H., McLellan, M., Yau, C., Kandoth, C., Bowlby, R., Shen, H., Hayat, S., Fieldhouse, R., Lester, S.C., Tse, G.M., Factor, R.E., Collins, L.C., Allison, K.H., Chen, Y.Y., Jensen, K., Johnson, N.B., Oesterreich, S., Mills, G.B., Cherniack, A.D., Robertson, G., Benz, C., Sander, C., Laird, P.W., Hoadley, K.A., King, T.A., Perou, C.M., 2015. Comprehensive molecular portraits of invasive lobular breast cancer. Cell 163, 506–519.

Cui, Y., Guadagno, T., 2008. B-RafV600E signaling deregulates the mitotic spindle checkpoint through stabilizing Mps1 levels in melanoma cells. Oncogene 27, 3122–3133.

Daniel, J., Coulter, J., Woo, J.-H., Wilsbach, K., Gabrielson, E., 2011. High levels of the Mps1 checkpoint protein are protective of aneuploidy in breast cancer cells. Proc. Natl. Acad. Sci. 108, 5384–5389.

Diallo, R., Ting, E., Gluz, O., Herr, A., Schütt, G., Geddert, H., Mohrmann, S., Gabbert, H., Nitz, U., Poremba, C., 2006. C-kit expression in high-risk breast cancer subgroup treated with high-dose or conventional dose-dense chemotherapy. Verh. Dtsch. Ges. Pathol. 90, 177–185.

ELIYATKıN, N., YALÇıN, E., ZENGEL, B., AKTAŞ, S., VARDAR, E., 2015. Molecular classification of breast carcinoma: from traditional, old-fashioned way to a new age, and a new way. J. Breast Health 11, 59.

Ellsworth, R.E., Seebach, J., Field, L.A., Heckman, C., Kane, J., Hooke, J.A., Love, B., Shriver, C.D., 2009. A gene expression signature that defines breast cancer metastases. Clin. Exp. Metastasis 26, 205–213.

Ergin, S., Kherad, N., Alagoz, M., 2022. RNA sequencing and its applications in cancer and rare diseases. Mol. Biol. Rep., 1–9

Faramarzi, A., Jahromi, M.G., Ashourzadeh, S., Jalilian, N., 2021. Metastatic and pathophysiological characteristics of breast cancer with emphasis on hereditary factors. Central Asian J. Med. Pharma. Sci. Innov 1, 104–113.

Feng, Y., Sun, B., Li, X., Zhang, L., Niu, Y., Xiao, C., Ning, L., Fang, Z., Wang, Y., Zhang, L., 2007. Differentially expressed genes between primary cancer and paired lymph node metastases predict clinical outcome of node-positive breast cancer patients. Breast Cancer Res. Treat. 103, 319–329.

Gachechiladze, M., Škarda, J., Soltermann, A., Joerger, M., 2017. RAD51 as a potential surrogate marker for DNA repair capacity in solid malignancies. Int. J. Cancer 141, 1286–1294.

Gordon, L.A., Mulligan, K.T., Maxwell-Jones, H., Adams, M., Walker, R.A., Jones, J.L., 2003. Breast cell invasive potential relates to the myoepithelial phenotype. Int. J. Cancer 106, 8–16.

Grundy, M.K., Buckanovich, R.J., Bernstein, K.A., 2020. Regulation and pharmacological targeting of RAD51 in cancer. Nar Cancer 2, zcaa024.

Hao, X., Sun, B., Hu, L., Lähdesmäki, H., Dunmire, V., Feng, Y., Zhang, S.W., Wang, H., Wu, C., Wang, H., 2004. Differential gene and protein expression in primary breast malignancies and their lymph node metastases as revealed by combined cDNA microarray and tissue microarray analysis. Cancer: Interdiscip. Int. J. Am. Cancer Soc. 100, 1110–1122.

HERNÁNDEZ-ROJAS, R., JIMÉNEZ-ARELLANO, C., DE LA FUENTE-GRANADA, M., ORDAZ-ROSADO, D., GARCÍA-BECERRA, R., VALENCIA-MAYORAL, P., DE LOURDES ÁLVAREZ-ARELLANO, M., EGUÍA-AGUILAR, P., VELASCO-VELÁZQUEZ, M. A. & GONZÁLEZ-ARENAS, A. 2022. The interplay between estrogen receptor beta and protein kinase C, a crucial collaboration for medulloblastoma cell proliferation and invasion. *Cellular Signalling***,** 110246

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1–13.

Hwang, B., Lee, J.H., Bang, D., 2018. Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp. Mol. Med. 50, 1–14.

Ikemura, S., Aramaki, N., Fujii, S., Kirita, K., Umemura, S., Matsumoto, S., Yoh, K., Niho, S., Ohmatsu, H., Kuwata, T., 2017. Changes in the tumor microenvironment during lymphatic metastasis of lung squamous cell carcinoma. Cancer Sci. 108, 136–142.

Inano, S., Sato, K., Katsuki, Y., Kobayashi, W., Tanaka, H., Nakajima, K., Nakada, S., Miyoshi, H., Knies, K., Takaori-Kondo, A., 2017. RFWD3-mediated ubiquitination promotes timely removal of both RPA and RAD51 from DNA damage sites to facilitate homologous recombination. Mol. Cell 66, (622–634) e8.

Janostiak, R., Vyas, M., Cicek, A.F., Wajapeyee, N., Harigopal, M., 2018. Loss of c-KIT expression in breast cancer correlates with malignant transformation of breast epithelium and is mediated by KIT gene promoter DNA hypermethylation. Exp. Mol. Pathol. 105, 41–49.

JEMAL, A., SIEGEL, R., WARD, E., HAO, Y., XU, J. & THUN, M. J. 2009. Cancer statistics, 2009. *CA: a cancer journal for clinicians,* 59**,** 225-249.

Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., Bork, P., von Mering, C., 2009. STRING 8–a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 37, D412–D416.

Jiao, M., Zhang, F., Teng, W., Zhou, C., 2022. MYBL2 is a novel independent prognostic biomarker and correlated with immune infiltrates in prostate cancer. Int. J. General Med. 15, 3003.

Karimnia, N., Ho, G.Y., Stephens, A.N., Bilandzic, M., 2021. Targeting Leader Cells in Ovarian Cancer as an Effective Therapeutic Option. Ovarian Cancer-Updates in Tumour Biology and Therapeutics, IntechOpen.

Kashiwagi, S., Yashiro, M., Takashima, T., Aomatsu, N., Kawajiri, H., Ogawa, Y., Onoda, N., Ishikawa, T., Wakasa, K., Hirakawa, K., 2013. c-Kit expression as a prognostic molecular marker in patients with basal-like breast cancer. J. Brit. Surg. 100, 490–496.

Kim, H.-S., Park, I., Cho, H.J., Gwak, G., Yang, K., Bae, B.N., Kim, K.W., Han, S., Kim, H.-J., Kim, Y.-D., 2012. Analysis of the potent prognostic factors in luminal-type breast cancer. J. Breast Cancer 15, 401–406.

KOBOLDT, D. C., FULTON, R. S., MCLELLAN, M. D., SCHMIDT, H., KALICKI-VEIZER, J., MCMICHAEL, J. F., FULTON, L. L., DOOLING, D. J., DING, L., MARDIS, E. R., WILSON, R. K., ALLY, A., BALASUNDARAM, M., BUTTERFIELD, Y. S. N., CARLSEN, R., CARTER, C., CHU, A., CHUAH, E., CHUN, H.-J. E., COOPE, R. J. N., DHALLA, N., GUIN, R., HIRST, C., HIRST, M., HOLT, R. A., LEE, D., LI, H. I., MAYO, M., MOORE, R. A., MUNGALL, A. J., PLEASANCE, E., GORDON ROBERTSON, A., SCHEIN, J. E., SHAFIEI, A., SIPAHIMALANI, P., SLOBODAN, J. R., STOLL, D., TAM, A., THIESSEN, N., VARHOL, R. J., WYE, N., ZENG, T., ZHAO, Y., BIROL, I., JONES, S. J. M., MARRA, M. A., CHERNIACK, A. D., SAKSENA, G., ONOFRIO, R. C., PHO, N. H., CARTER, S. L., SCHUMACHER, S. E., TABAK, B., HERNANDEZ, B., GENTRY, J., NGUYEN, H., CRENSHAW, A., ARDLIE, K., BEROUKHIM, R., WINCKLER, W., GETZ, G., GABRIEL, S. B., MEYERSON, M., CHIN, L., PARK, P. J., KUCHERLAPATI, R., HOADLEY, K. A., TODD AUMAN, J., FAN, C., TURMAN, Y. J., SHI, Y., LI, L., TOPAL, M. D., HE, X., CHAO, H.-H., PRAT, A., SILVA, G. O., IGLESIA, M. D., ZHAO, W., USARY, J., BERG, J.

S., ADAMS, M., BOOKER, J., WU, J., GULABANI, A., BODENHEIMER, T., HOYLE, A. P., SIMONS, J. V., SOLOWAY, M. G., MOSE, L. E., JEFFERYS, S. R., BALU, S., PARKER, J. S., NEIL HAYES, D., PEROU, C. M., MALIK, S., MAHURKAR, S., SHEN, H., WEISENBERGER, D. J., TRICHE JR, T., et al. 2012. Comprehensive molecular portraits of human breast tumours. *Nature,* 490**,** 61-70

Kremer, L.S., Bader, D.M., Mertes, C., Kopajtich, R., Pichler, G., Iuso, A., Haack, T.B., Graf, E., Schwarzmayr, T., Terrile, C., 2017. Genetic diagnosis of Mendelian disorders via RNA sequencing. Nat. Commun. 8, 1–11.

Kumar, R., Sharma, A., Tiwari, R.K., 2012. Application of microarray in breast cancer: an overview. J. Pharma. Bioall. Sci. 4, 21–26.

Lichtner, R., Julian, J., North, S., Glasser, S., Nicolson, G., 1991. Coexpression of cytokeratins characteristic for myoepithelial and luminal cell lineages in rat 13762NF mammary adenocarcinoma tumors and their spontaneous metastases. Cancer Res. 51, 5943–5950.

LIGHTBODY, G., HABERLAND, V., BROWNE, F., TAGGART, L., ZHENG, H., PARKES, E. & BLAYNEY, J. K. 2019. Review of applications of high-throughput sequencing in personalized medicine: barriers and facilitators of future progress in research and clinical application. Briefings in bioinformatics, 20, 1795-1811.

Lin, G., Chai, J., Yuan, S., Mai, C., Cai, L., Murphy, R.W., Zhou, W., Luo, J., 2016. VennPainter: a tool for the comparison and identification of candidate genes based on venn diagrams. PLoS ONE 11, e0154315.

Linde, N., Casanova-Acebes, M., Sosa, M.S., Mortha, A., Rahman, A., Farias, E., Harper, K., Tardio, E., Reyes torres, i. & jones, j., 2018. Macrophages orchestrate breast cancer early dissemination and metastasis. Nat. Commun. 9, 1–14.

Liu, S., Chen, X., Wang, Y., Li, L., Wang, G., Li, X., Chen, H., Guo, J., Lin, H., Lian, Q.-Q., Ge, R.-S., 2017. A role of KIT receptor signaling for proliferation and differentiation of rat stem Leydig cells in vitro. Mol. Cell. Endocrinol. 444, 1–8.

Liu, X., Liao, W., Yuan, Q., Ou, Y., Huang, J., 2015. TTK activates Akt and promotes proliferation and migration of hepatocellular carcinoma cells. Oncotarget 6, 34309.

Liu, Y., Zhu, K., Guan, X., Xie, S., Wang, Y., Tong, Y., Guo, L., Zheng, H., Lu, R., 2021. TTK is a potential therapeutic target for cisplatin-resistant ovarian cancer. J. Ovarian Res. 14, 1–10.

Liu, Z., Guo, Z., Long, L., Zhang, Y., Lu, Y., Wu, D., Dong, Z., 2020. Spindle assembly checkpoint complex-related genes TTK and MAD2L1 are over-expressed in lung adenocarcinoma: a big data and bioinformatics analysis. Nan Fang yi ke da xue xue bao= J. Southern Med. Univ. 40, 1422–1431.

Loughman, T., Barron, S., Wang, C.-J., Dynoodt, P., Fender, B., Lopez-Ruiz, C., Stapleton, S., Fabre, A., Quinn, C., Nodin, B., Jirström, K., Razmara, F., O'Grady, A., Baird, A.-M., Gray, S.G., Freixo, A., Moelans, C.B., van Diest, P.J., Duffy, M.J., O'Leary, D., Crown, J., Bracken, A.P., Gallagher, W.M., 2022. Analytical validation of a novel 6-gene signature for prediction of distant recurrence in estrogen receptor-positive, HER2-negative, early-stage breast cancer. Clin. Chem.

Luo, S., Cao, N., Tang, Y., Gu, W., 2017. Identification of key microRNAs and genes in preeclampsia by bioinformatics analysis. PLoS ONE 12, e0178549.

Ly, A., Lester, S.C., Dillon, D., 2012. Prognostic factors for patients with breast cancer: traditional and new. Surg. Pathol. Clin. 5, 775–785.

Maacke, H., Opitz, S., Jost, K., Hamdorf, W., Henning, W., Krüger, S., Feller, A.C., Lopens, A., Diedrich, K., Schwinger, E., 2000. Over-expression of wild-type Rad51 correlates with histological grading of invasive ductal breast cancer. Int. J. Cancer 88, 907–913.

Maire, V., Baldeyron, C., Richardson, M., Tesson, B., Vincent-Salomon, A., Gravier, E., Marty-Prouvost, B., de Koning, L., Rigaill, G., Dumont, A., 2013. TTK/hMPS1 is an attractive therapeutic target for triple-negative breast cancer. PLoS ONE 8, e63712.

MAITRA, D. & SRIVASTAVA, A. 2022. Tumour Markers, Prognostic and Predictive Factors in Breast Cancer. *Breast Cancer.* Springer.

Marco-Puche, G., Lois, S., Benítez, J., Trivino, J.C., 2019. RNA-Seq perspectives to improve clinical diagnosis. Front. Genet. 10, 1152.

Marsh, T., Debnath, J., 2020. Autophagy suppresses breast cancer metastasis by degrading NBR1. Autophagy 16, 1164–1165.

Masood, S., 2005. Prognostic/predictive factors in breast cancer. Clin. Lab. Med. 25, 809–825.

Mimori, K., Kataoka, A., Yoshinaga, K., Ohta, M., Sagara, Y., Yoshikawa, Y., Ohno, S., Barnard, G.F., Mori, M., 2005. Identification of molecular markers for metastasis-related genes in primary breast cancer cells. Clin. Exp. Metastasis 22, 59–67.

Montor, W.R., Salas, A.R.O.S.E., Melo, F.H.M.D., 2018. Receptor tyrosine kinases and downstream pathways as druggable targets for cancer treatment: the current arsenal of inhibitors. Mol. Cancer 17, 1–18.

Nathanson, S.D., 2003. Insights into the mechanisms of lymph node metastasis. Cancer 98, 413–423.

Nazarov, P.V., Muller, A., Kaoma, T., Nicot, N., Maximo, C., Birembaut, P., Tran, N.L., Dittmar, G., Vallar, L., 2017. RNA sequencing and transcriptome arrays analyses show opposing results for alternative splicing in patient derived samples. BMC Genomics 18, 1–18.

Negi, A., Shukla, A., Jaiswar, A., Shrinet, J., Jasrotia, R.S., 2022. Applications and challenges of microarray and RNA-sequencing. In: Bioinformatics. Elsevier, pp. 91–103. https://doi.org/10.1016/B978-0-323-89775-4.00016-X.

OZKAN, S. 2022. Psychosocial aspects of breast cancer. *Global Perspectives in Cancer Care: Religion, Spirituality, and Cultural Diversity in Health and Healing*, 302.

Pachis, S.T., Kops, G.J.P.L., 2018. Leader of the SAC: molecular mechanisms of Mps1/TTK regulation in mitosis. Open Biol. 8, 180109.

Pachmayr, E., Treese, C., Stein, U., 2017. Underlying mechanisms for distant metastasis-molecular biology. Visceral Med. 33, 11–20.

PANEL, N. I. O. H. C. D. 2001. National Institutes of Health Consensus Development Conference statement: adjuvant therapy for breast cancer, November 1–3, 2000. *JNCI Monographs,* 2001**,** 5-15

Pangou, E., Sumara, I., 2021. The multifaceted regulation of mitochondrial dynamics during mitosis. Front. Cell Dev. Biol. 9.

Papafotiou, G., Paraskevopoulou, V., Vasilaki, E., Kanaki, Z., Paschalidis, N., Klinakis, A., 2016. KRT14 marks a subpopulation of bladder basal cells with pivotal role in regeneration and tumorigenesis. Nat. Commun. 7, 1–11.

Pearson, G.W., 2019. Control of invasion by epithelial-to-mesenchymal transition programs during metastasis. J. Clin. Med. 8, 646.

Petrocca, F., Altschuler, G., Tan, S., Mendillo, M., Yan, H., Jerry, D.J., Kung, A., Hide, W., Ince, T., Lieberman, J., 2013. A genome-wide siRNA screen identifies proteasome addiction as a vulnerability of basal-like triple-negative breast cancer cells. Cancer Cell 24 (2), 182–196.

Prakash, R., Zhang, Y., Feng, W., Jasin, M., 2015. Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins. Cold Spring Harbor Perspect. Biol. 7, a016600.

Raderschall, E., Stout, K., Freier, S., Suckow, V., Schweiger, S., Haaf, T., 2002. Elevated levels of Rad51 recombination protein in tumor cells. Cancer Res. 62, 219–225.

Ren, S., Peng, Z., Mao, J.-H., Yu, Y., Yin, C., Gao, X., Cui, Z., Zhang, J., Yi, K., Xu, W., Chen, C., Wang, F., Guo, X., Lu, J.i., Yang, J., Wei, M., Tian, Z., Guan, Y., Tang, L., Xu, C., Wang, L., Gao, X.u., Tian, W., Wang, J., Yang, H., Wang, J., Sun, Y., 2012. RNA-seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions, cancer-associated long noncoding RNAs and aberrant alternative splicings. Cell Res. 22 (5), 806–821.

Rock, J.R., Onaitis, M.W., Rawlins, E.L., Lu, Y., Clark, C.P., Xue, Y., Randell, S.H., Hogan, B.L., 2009. Basal cells as stem cells of the mouse trachea and human airway epithelium. Proc. Natl. Acad. Sci. 106, 12771–12775.

Rogers, C., Loveday, R., Drew, P., Greenman, J., 2002. Molecular prognostic indicators in breast cancer. Eur. J. Surg. Oncol. (EJSO) 28, 467–478.

ROYCHOWDHURY, S. & CHINNAIYAN, A. M. 2016. Translating cancer genomes and transcriptomes for precision oncology. *CA: a cancer journal for clinicians,* 66**,** 75-88.

Salto-tellez, m. & gonzalez de castro, d., 2014. Next-generation sequencing: a change of paradigm in molecular diagnostic validation. J. Pathol. 234, 5–10.

Salvatore, G., Nappi, T.C., Salerno, P., Jiang, Y., Garbi, C., Ugolini, C., Miccoli, P., Basolo, F., Castellone, M.D., Cirafici, A.M., 2007. A cell proliferation and chromosomal instability signature in anaplastic thyroid carcinoma. Cancer Res. 67, 10148–10158.

Sauerbrei, W., Royston, P., Bojar, H., Schmoor, C., Schumacher, M., 1999. Modelling the effects of standard prognostic factors in node-positive breast cancer. Br. J. Cancer 79, 1752–1760.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504.

Sherman, B.T., Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57.

Stark, R., Grzelak, M., Hadfield, J., 2019. RNA sequencing: the teenage years. Nat. Rev. Genet. 20, 631–656.

Sullivan, M.R., Bernstein, K.A., 2018. RAD-ical new insights into RAD51 regulation. Genes 9, 629.

Suyal, G., Pandey, P., Saraya, A., Sharma, R., 2022. Tumour suppressor role of microRNA-335-5p in esophageal squamous cell carcinoma by targeting TTK (Mps1). Exp. Mol. Pathol. 124, 104738.

Tang, Z., Li, C., Kang, B., Gao, G., Li, C., Zhang, Z., 2017. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 45, W98–W102.

Tannous, B.A., Kerami, M., van der Stoop, P.M., Kwiatkowski, N., Wang, J., Zhou, W., Kessler, A.F., Lewandrowski, G., Hiddingh, L., Sol, N., 2013. Effects of the selective MPS1 inhibitor MPS1-IN-3 on glioblastoma sensitivity to antimitotic drugs. J. Natl Cancer Inst. 105, 1322–1331.

Theelen, W.S.M.E., Krijgsman, O., Monkhorst, K., Kuilman, T., Peters, D.D.G.C., Cornelissen, S., Ligtenberg, M.A., Willems, S.M., Blaauwgeers, J.L.G., van Noesel, C.J.M., Peeper, D.S., van den Heuvel, M.M., Schulze, K., 2020. Presence of a 34-gene signature is a favorable prognostic marker in squamous non-small cell lung carcinoma. J. Transl. Med. 18, 271.

Vecchi, M., Confalonieri, S., Nuciforo, P., Vigano, M., Capra, M., Bianchi, M., Nicosia, D., Bianchi, F., Galimberti, V., Viale, G., 2008. Breast cancer metastases are molecularly distinct from their primary tumors. Oncogene 27, 2148–2158.

Volkmer, J.-P., Sahoo, D., Chin, R.K., Ho, P.L., Tang, C., Kurtova, A.V., Willingham, S.B., Pazhanisamy, S.K., Contreras-Trujillo, H., Storm, T.A., 2012. Three differentiation states risk-stratify bladder cancer into distinct subtypes. Proc. Natl. Acad. Sci. 109, 2078–2083.

Wang, J., Xie, Y., Bai, X., Wang, N., Yu, H., Deng, Z., Lian, M., Yu, S., Liu, H., Xie, W., 2018. Targeting dual specificity protein kinase TTK attenuates tumorigenesis of glioblastoma. Oncotarget 9, 3081.

Ward, A., Balwierz, A., Zhang, J.D., Küblbeck, M., Pawitan, Y., Hielscher, T., Wiemann, S., Sahin, Ö., 2013. Re-expression of microRNA-375 reverses both tamoxifen resistance and accompanying EMT-like properties in breast cancer. Oncogene 32, 1173–1182.

Werner, S., Keller, L., Pantel, K., 2020. Epithelial keratins: biology and implications as diagnostic markers for liquid biopsies. Mol. Aspects Med. 72, 100817.

WINTHEISER, G. A. & SILBERSTEIN, P. 2021. Physiology, tyrosine kinase receptors. *StatPearls [Internet].* StatPearls Publishing

Xie, Y., Wang, A., Lin, J., Wu, L., Zhang, H., Yang, X., Wan, X., Miao, R., Sang, X., Zhao, H., 2017. Mps1/TTK: a novel target and biomarker for cancer. J. Drug Target. 25, 112–118.

Xu, Q., Xu, Y., Pan, B., Wu, L., Ren, X., Zhou, Y., Mao, F., Lin, Y., Guan, J., Shen, S., Zhang, X., Wang, C., Zhong, Y., Zhou, L., Liang, Z., Zhao, H., Sun, Q., 2016. TTK is a favorable prognostic biomarker for triple-negative breast cancer survival. Oncotarget 7, 81815–81829.

Xu, Y.-H., Deng, J.-L., Wang, L.-P., Zhang, H.-B., Tang, L., Huang, Y., Tang, J., Wang, S.-M., Wang, G., 2020. Identification of candidate genes associated with breast cancer prognosis. DNA Cell Biol. 39, 1205–1227.

Yao, Z.-P., Zhu, H., Shen, F., Gong, D., 2021. Hsp90 regulates the tumorigenic function of tyrosine protein kinase in osteosarcoma. Clin. Exp. Pharmacol. Physiol. 49 (3), 380–390.

Yao, Z.P., Zhu, H., Shen, F., Gong, D., 2022. Hsp90 regulates the tumorigenic function of tyrosine protein kinase in osteosarcoma. Clin. Exp. Pharmacol. Physiol. 49, 380–390.

Zhang, S., Cao, H., Ye, L., Wen, X., Wang, S., Zheng, W., Zhang, Y., Huang, D., Gao, Y., Liu, H., 2019a. Cancer-associated methylated lncRNAs in patients with bladder cancer. Am. J. Transl. Res. 11, 3790.

Zhang, X., Ma, N., Yao, W., Li, S., Ren, Z., 2019b. RAD51 is a potential marker for prognosis and regulates cell proliferation in pancreatic cancer. Cancer Cell Int. 19, 1–11.

Zhou, Z., Zhang, C., Ma, Z., Wang, H., Tuo, B., Cheng, X., Liu, X., Li, T., 2022. Pathophysiological role of ion channels and transporters in HER2-positive breast cancer. Cancer Gene Ther., 1–8