

RESEARCH

Open Access



# Meta-analysis of integrated ChIP-seq and transcriptome data revealed genomic regions affected by estrogen receptor alpha in breast cancer

Zeynab Piryaei<sup>1,2</sup>, Zahra Salehi<sup>2,3</sup>, Esmail Ebrahimie<sup>4</sup>, Mansour Ebrahimi<sup>5</sup> and Kaveh Kavousi<sup>2\*</sup>

## Abstract

**Background** The largest group of patients with breast cancer are estrogen receptor-positive (ER<sup>+</sup>) type. The estrogen receptor acts as a transcription factor and triggers cell proliferation and differentiation. Hence, investigating ER-DNA interaction genomic regions can help identify genes directly regulated by ER and understand the mechanism of ER action in cancer progression.

**Methods** In the present study, we employed a workflow to do a *meta-analysis* of ChIP-seq data of ER<sup>+</sup> cell lines stimulated with 10 nM and 100 nM of E2. All publicly available data sets were re-analyzed with the same platform. Then, the known and unknown batch effects were removed. Finally, the *meta-analysis* was performed to obtain meta-differentially bound sites in estrogen-treated MCF7 cell lines compared to vehicles (as control). Also, the *meta-analysis* results were compared with the results of T47D cell lines for more precision. Enrichment analyses were also employed to find the functional importance of common meta-differentially bound sites and associated genes among both cell lines.

**Results** Remarkably, *POU5F1B*, *ZNF662*, *ZNF442*, *KIN*, *ZNF410*, and *SGSM2* transcription factors were recognized in the *meta-analysis* but not in individual studies. Enrichment of the meta-differentially bound sites resulted in the candidacy of pathways not previously reported in breast cancer. *PCGF2*, *HNF1B*, and *ZBED6* transcription factors were also predicted through the enrichment analysis of associated genes. In addition, comparing the *meta-analysis* results of both ChIP-seq and RNA-seq data showed that many transcription factors affected by ER were up-regulated.

**Conclusion** The *meta-analysis* of ChIP-seq data of estrogen-treated MCF7 cell line leads to the identification of new binding sites of ER that have not been previously reported. Also, enrichment of the meta-differentially bound sites and their associated genes revealed new terms and pathways involved in the development of breast cancer which should be examined in future in vitro and in vivo studies.

**Keywords** Meta-analysis, ChIP-seq, RNA-seq, Breast cancer, Estrogen receptor-positive, MCF7/T47D cell lines

\*Correspondence:

Kaveh Kavousi

kkavousi@ut.ac.ir

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Introduction

Breast cancer (BC) is the world's most widespread cancer among women [1]. The majority of BC patients are estrogen receptor-positive (ER<sup>+</sup>), i.e., Cancer cells have estrogen receptors (ESR1) [2, 3]. When the ER interacts with estrogen, it acts as a transcription factor (TF) and triggers cell proliferation and differentiation [4]. Hence, finding ER-DNA interaction genomic regions can help identify genes that are regulated by ER. Specifically, investigating these genomic regions can help to understand ER modes of action in cancer progression.

Chromatin immunoprecipitation sequencing (ChIP-seq), as a next-generation sequencing technique, is applied to find the transcription factor binding sites (TFBSs) [5]. In the present era, several ChIP-seq datasets have been generated to find differentially bound sites (DBSs) in BC patients [6–8]. Each study has results unique or commonalities with other related studies, which could be due to known and unknown batch effects [6–8]. On the other hand, gathering relevant studies and performing *meta-analysis* supplies more exact results [9]. Few studies have been performed on the *meta-analysis* of ChIP-seq data. A *meta-analysis* study was conducted by Kolmykov and colleagues on ChIP-seq datasets through the rank aggregation approach, and the significant TFBSs in BC were identified [10]. Some complexities with ChIP-seq data should be considered for *meta-analysis*, including: (1) it is necessary to consider the same cell line, dose, and treatment period (known batch effects), (2) selecting the appropriate treatment period, i.e., after which treatment period, treated and untreated samples show the most differentiation, (3) removing the effects of differences laboratory conditions (unknown batch effects) seems to be vitally important to obtain more precise *meta-analysis* results, (4) since each study may have been analyzed with different platforms and genomic references, it is necessary to re-analyze the samples with the same tools and references, and (5) it is noteworthy that even if the same platform is applied, a different number of peaks and regions are obtained for each sample. Therefore after combining the datasets, they must be prepared to obtain the same number, regions, and scores of TFBSs for all samples. Next, the unknown batch effects must be removed before *meta-analysis*. Despite the necessities mentioned above, no *meta-analysis* study has already been performed on integrated TFBSs to obtain significant binding sites in BC.

In the current study, to obtain meta-differentially bound sites in estrogen receptor-positive breast cancer cell lines, we employed an innovative workflow for the meta-analysis of ChIP-seq data. Given the ER's essential role in BC, MCF7 and T47D cell lines stimulated with E2 were collected to find genomic regions directly

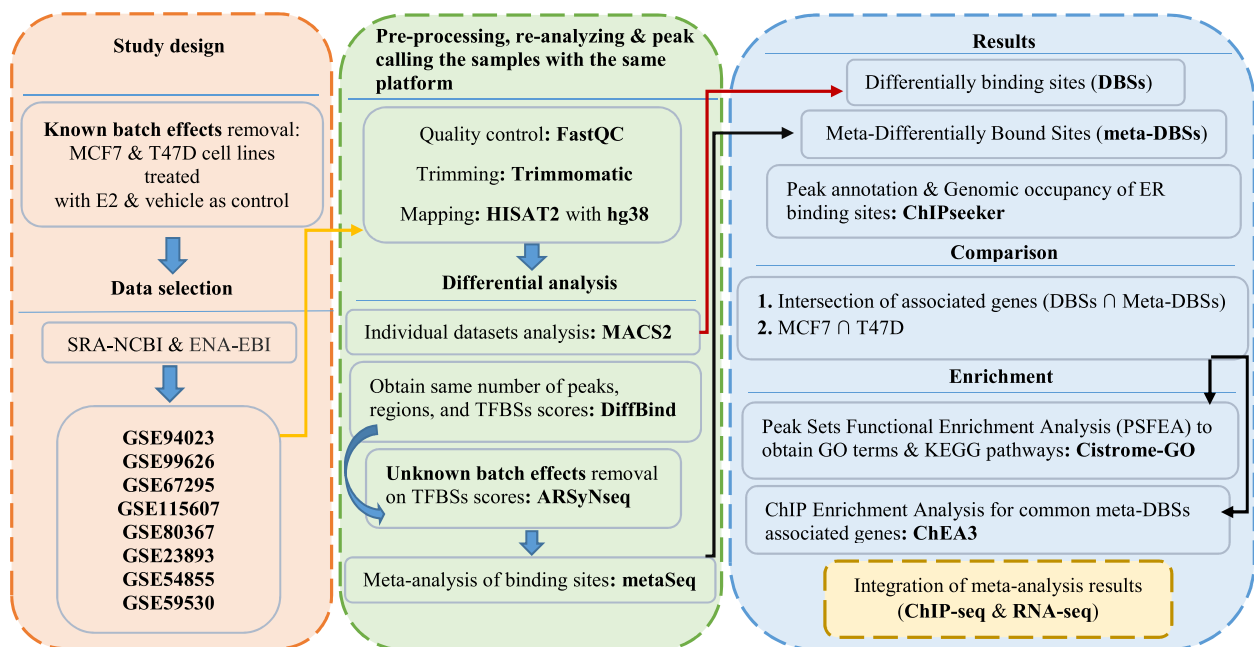
interacting with ER. MCF7 and T47D cell lines, representing luminal A subtype, are extensively utilized as experimental models in breast cancer research, particularly for the study of hormone-dependent BC. These two ER+BC cell lines serve as valuable tools in understanding the molecular mechanisms and exploring potential therapeutic interventions for BC, particularly luminal A subtype. In the MCF7 cell line, concerning the same cell line, dose, and treatment period (removing known batch effects), public ChIP-seq datasets were selected and re-analyzed. Next, samples were integrated, and the same number, regions, and scores of TFBSs were obtained in doses of 10 nM and 100 nM separately. Finally, unknown batch effects were removed, and several ChIP-seq datasets were meta-analyzed based on TFBSs scores. Hence, we take advantage of the term meta-differentially bound sites (meta-DBSs) for the first time. For more precision, the *meta-analysis* results were compared with the results of T47D cell lines, and the intersection of both cell lines was obtained. By enriching the shared meta-DBSs and their associated genes, TFs and pathways that had not been previously reported in BC were identified. Moreover, the results of the ChIP-seq data *meta-analysis* were confirmed by comparing the results of the RNA-seq data *meta-analysis*.

## Materials and methods

After collecting data from SRA-NCBI [11] and ENA-EBI [12] databases, nine MCF7 and T47D cell line-associated ER alpha-ChIP-seq datasets from eight studies were selected. Differential analysis for each dataset was performed, and DBSs were obtained following data preprocessing and mapping. Next, the *meta-analysis* was performed, and meta-DBSs were identified in the MCF7 cell line treated with 10 nM and 100 nM E2 separately. Then, Peak Annotation was applied for DBSs and meta-DBSs, and the results of both cell lines were compared. Peak set functional enrichment analysis (PSFEA) and ChIP enrichment analysis (ChEA) were implemented for the common meta-DBSs and their associated genes, respectively. Also, the *meta-analysis* results of both ChIP-seq and RNA-seq data were compared. The general steps of the present study are summarized in Fig. 1.

## Data collection

A comprehensive search was conducted on SRA-NCBI [11] and ENA-EBI [12] databases to find ChIP-seq datasets in BC (Fig. 2). Some criteria were considered in the selection of the ChIP-seq datasets: (1) association with ER<sup>+</sup> cell lines; MCF7 and T47D cell lines, (2) availability of both treated with vehicles (as controls) and treated with E2 samples, (3) datasets without any knocked out genes, and (4) datasets with the same and appropriate



**Fig. 1** An overview of the *meta-analysis* steps performed in the present study. ChIP-seq datasets were retrieved from SRA-NCBI and ENA-EBI databases. Pre-processing and re-analyzing steps of datasets were conducted. Then, batch effects removal and *meta-analysis* were performed. Subsequently, peak set functional enrichment analysis (PSFEA) and ChIP Enrichment Analysis for meta-DBSs-associated common genes were performed using the Cistrome-GO database and ChEA3 database, respectively. The packages and methods employed are in bold form. *TFBSs* Transcription factor binding sites, *DBSs* Differentially bound sites, *Meta-DBSs* Meta-differentially bound sites

dose and period of treatment. Consequently, GSE94023, GSE99626, GSE67295, GSE115607, GSE80367, GSE23893, GSE54855, and GSE59530, including MCF7 and T47D cell lines stimulated with 10 nM and 100 nM E2 for 40 or 45 min were selected. The details of the datasets are described in Table 1.

### Pre-processing and data analysis

The selected ChIP-seq datasets were re-analyzed under the Galaxy platform (<https://galaxyproject.org>) [13] to homogenize studies. The data quality control and trimming were performed using FastQC (version 0.11.5) [14] and Trimmomatic (version 0.38) [15], respectively. The human reference genome (hg38) was utilized for data mapping by HISAT2 (version 2.1.0) [16].

### Analysis of individual datasets to identify DBSs

Within each individual study, DBSs were identified in MCF7 and T47D cell lines treated with E2 compared to treated with vehicle (as control) through MACS2 [22] (version 2.1.1.20160309.6) ( $-\log_{10}(q\text{-value}) > 2$ ; fold-enrichment  $> 2$ ). Then, the ChIPseeker (version 1.26.2) [23] package was applied for peak annotation. The identified DBSs associated genes of individual studies were further utilized to compare with the results of the ChIP-seq *meta-analysis*. During peak calling with MACS2, the

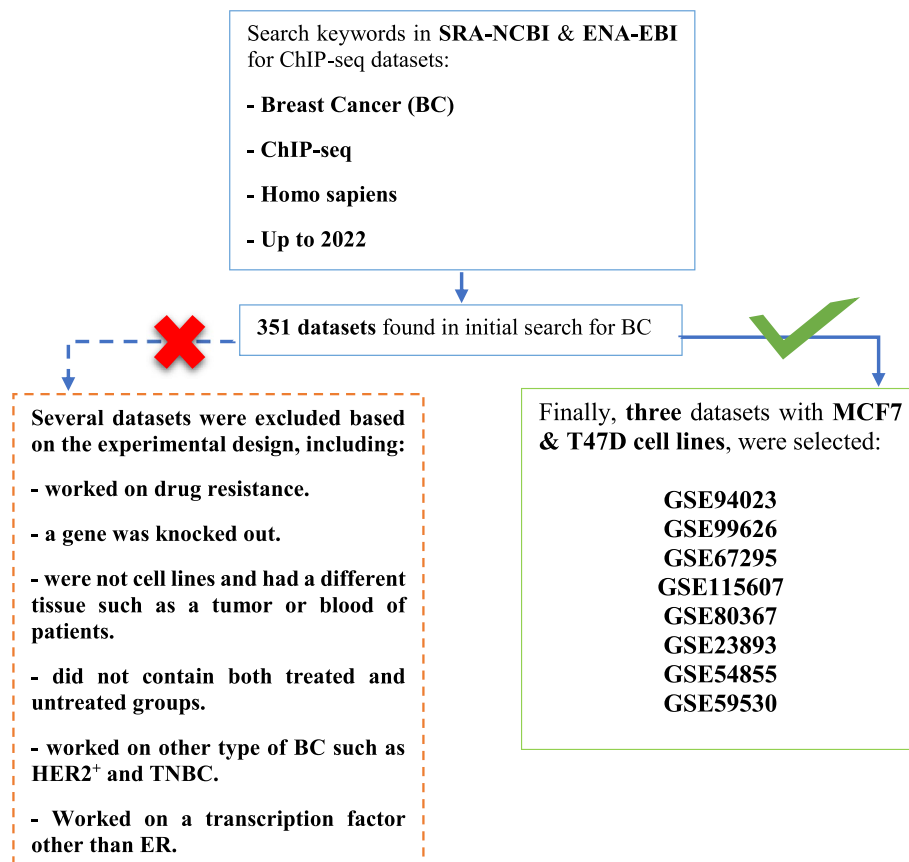
input samples were also considered to reduce noise. The Blacklist regions were also considered to improve the signal-to-noise ratio. The maximum tags were set to keep one tag at the same location ( $-\text{keep-dup}=1$ ). In the individual dataset analysis, replicates were combined using the rmpsc (Multiple Sample Peak Calling) package [24].

Typically, analysis of treatment, vehicle, and input samples on ChIP-seq data is performed in the following steps:

- A. ER-ChIP treated with E2 vs. input = corrected ER-ChIP treated with E2.
- B. Vehicle vs. input (for vehicle) = corrected vehicle.
- C. Corrected ER-ChIP treated with E2 vs. corrected Vehicle = detection of enriched regions in ER-ChIP treatment E2 (using MACS2).

### Meta-analysis to identify meta-DBSs

To find meta-DBSs, a *meta-analysis* workflow was utilized to compare cells of treated with vehicle versus treated with E2 at the TFBSs score levels (Fig. 3). First, the same number, regions, and scores of TFBSs were obtained utilizing the DiffBind package (Version 3.0.15) [25]. All replicates were considered as independent samples to produce the binding affinity matrix. Then,



**Fig. 2** The flowchart to select datasets. A total of 351 datasets from SRA-NCBI and ENA-EBI were evaluated. Finally, based on the four criteria described, eight studies on MCF7 and T47D cell lines treated with E2 were used in the present study

the ARSyNseq (ASCA Removal of Systematic Noise for sequencing data) method [26], implemented in the NOIseq package [26], was applied to remove unknown batch effects (Fig. 1). In ARSyNseq, the TMM (Trimmed Mean of M values) method [27] was utilized for between-sample normalization on TFBSs scores. Then, the metaSeq package [28] was employed on all samples to find meta-DBSs and set the statistical threshold as  $-\log_{10}(q\text{-value}) > 2$  to select meta-DBSs. Finally, the ChIPseeker package was utilized for peak annotation. All packages were implemented in R software. The parameters in the ARSyNseq and the metaSeq packages were set as follows:

```
mydata2corr1= ARSyNseq(mydata2, factor = "batch",
batch = TRUE, norm = "tmm")
result<- meta.oneside.noiseq(cds, k=0.5, norm = "n", replicates = "biological",
factor = flag1, conditions = c(1,0), studies = flag2)
```

#### Peak set functional enrichment analysis and ChIP enrichment analysis

Gene ontology (GO) and pathway enrichment analyses were performed with Cistrome-GO (the updated version of 2019) (<http://go.cistrome.org/>) [29] to

determine the functional importance of the identified meta-DBSs. For such analysis, we employed peak set functional enrichment analysis (PSFEA) term for the first time. The GO annotation was performed at three levels, including biological process (BP), molecular function (MF), and cellular component (CC). Furthermore, the Kyoto Encyclopedia of Genes and Genomes (KEGG) (the updated version of 2019) [30] enriched pathways were determined. The adjusted  $P$ -value  $\leq 0.05$  was applied to select significant GO terms and KEGG pathways. Also, a comprehensive ChIP Enrichment Analysis (ChEA) was performed on meta-DBSs-associated common genes using ChEA3 (the updated version of 2019) (<https://amp.pharm.mssm.edu/ChEA3>) [31], which contains six libraries, including Literature, Enrichr, ARCHS4—Coexpression, ENCODE, ReMap, and GTEx—Coexpression. Also, two overall results were obtained: (1) integrated\_topRank, and (2) integrated\_meanRank; in both of them TFs were ranked based on estimated scores in libraries. Adjusted  $P$ -value  $\leq 0.05$  was considered significant.



**Table 1** Characteristics of the selected ER-ChIP-seq datasets and their experimental design

Accession number	Cell lines	Dose (nM) of E2	Inputs	Samples treated with vehicles as controls	Samples treated with E2	chip antibody	Type of vehicle treatment	Experiment
GSE94023	MCF7	10	-	GSM2467220	GSM2467224	F3A6 (to ERalpha, Gronemeyer)	mock	Dzida, ... et al. [7]
GSE99626			GSM2125164 (Vehicle_Input) GSM2125165 (E2_Input)	GSM2648921	GSM2648922	anti-ER (SC biotech HC-20, cat# SC-543, lot# L1911)	ETOH	Singhal, ... et al. [17]
GSE67295			-	GSM1643946 (rep1) GSM1643947 (rep2)	GSM1643948 (rep1) GSM1643949 (rep2)	ERa (Vendor: Santa Cruz, cat# sc-543, lot# C2114)	DMSO	Stender, ... et al. [8]
GSE115607			GSM3184881	GSM3184871	GSM3184880	ERa (Santa Cruz, sc-543, lot F1716)	DMSO	Puyang, ... et al. [6]
GSE80367	T47D		GSM2112811 (Vehicle_Input) GSM2112812 (E2_Input)	GSM2112796	GSM2112798	anti-ER (SC biotech HC-20)	ETOH	Singhal, ... et al. [18]
GSE23893	T47D	100	GSM589247	GSM589238	GSM589239	ERa (sc-543, Santa-Cruz)	ETOH	Kong, ... et al. [19]
GSE23893	MCF7		GSM589244	GSM589236	GSM589237	ERa (sc-543, Santa-Cruz)	ETOH	Kong, ... et al. [19]
GSE54855			GSM1325252	GSM1325246	GSM1325250	ER (Santa Cruz, sc-542)	mock	Guertin, ... et al. [20]
GSE59530			GSM1534712 (Vehicle_Input_rep1) GSM1534713 (Vehicle_Input_rep2) GSM1534714 (E2_Input_rep1) GSM1534715 (E2_Input_rep2)	GSM1534720 (rep1) GSM1534721 (rep2)	GSM1534722 (rep1) GSM1534723 (rep2)	ER alpha (rabbit polyclonal generated in the Kraus Lab)	unknown	Franco, ... et al. [21]

nM Nanomolar, ETOH ethanol

## Results

The *meta-analysis* was performed on the ChIP-seq data based on TFBSs scores to identify meta-DBSs. Here, known batch effects were first removed. Since the treatment period with E2 can affect the results, several treatment periods were investigated to select the appropriately treated cell lines. In the GSE94023 study, there were various periods of treatment with E2, including 0 (mock treated as control) and 5, 10, 20, 40, 80, 160, 320, 640, and 1280 min (cases). An appropriate sampling time reveals the maximum distinction between treated and untreated samples. Hence, a heat map correlation matrix was drawn for all samples. According to the matrix, the samples after 40 min showed the most significant difference compared to the control (see Additional file 1: Figure S1). Thus, samples treated with

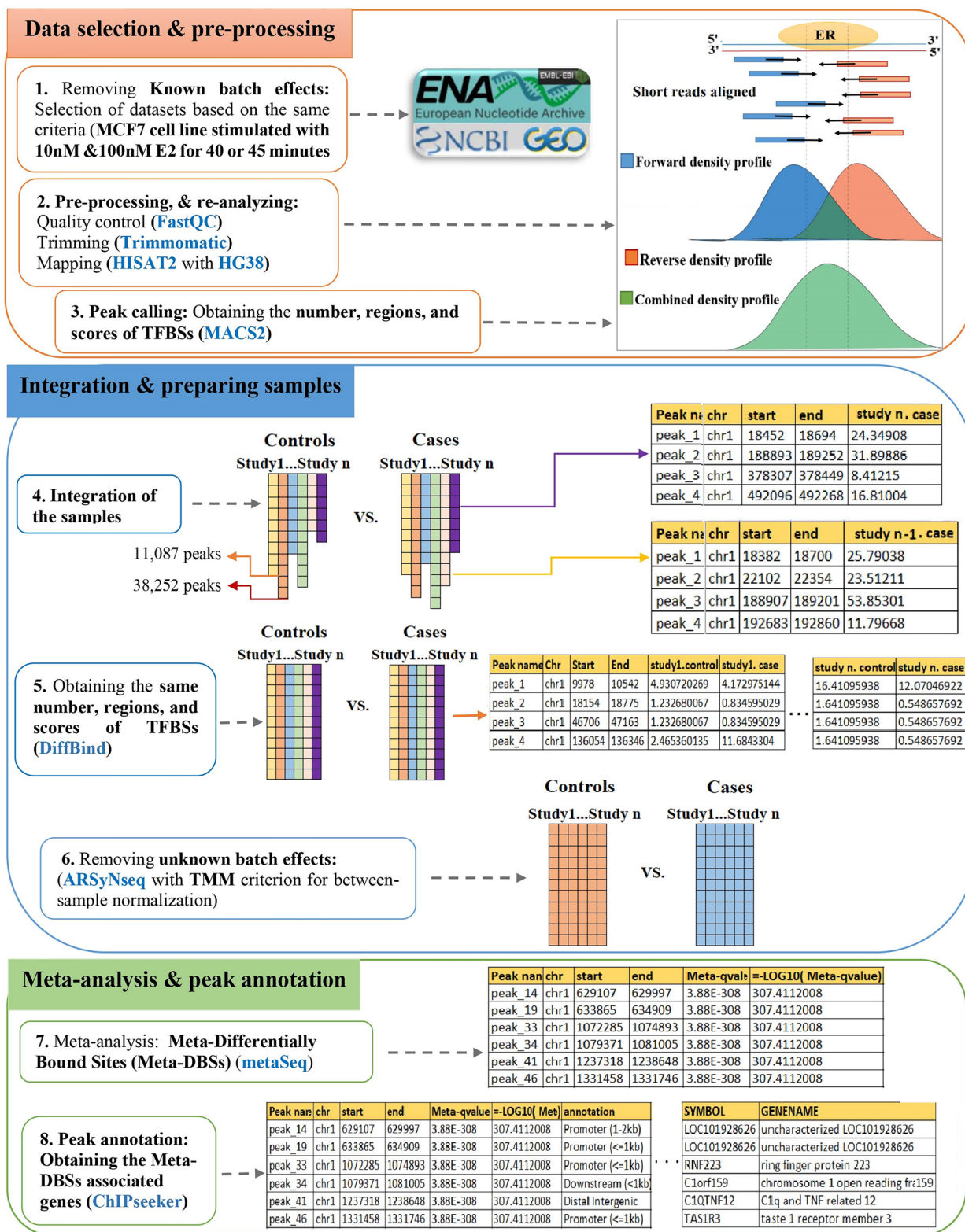
10 nM and 100 nM E2 for 40 or 45 min were selected. The minimum peak length was considered equal to 150-bp as the significant peak cutoff.

### Identification of DBSs

After selecting and pre-processing data, DBSs were obtained from nine individual ChIP-seq datasets (Table 2; see Additional file 2: Tables S1-S9) and compared with the *meta-analysis* results.

### Identification of meta-DBSs

To identify meta-DBSs in MCF7 cell lines, datasets were integrated based on stimulation with 10 nM or 100 nM E2 separately. Next, DiffBind was applied to obtain the same number, regions, and scores of TFBSs (see Additional file 2: Tables S10 and S11). Following unknown batch effect removal (see Additional file 1: Figures S2-S5),



**Table 2** Identified DBSs/ meta-DBSs through individual/ meta-analysis on eight studies of the MCF7 and T47D cell lines treated with 10 nM and 100 nM E2

Analysis	Accession number	Cell line	Treatment	Dose (nM)	DBSs or Meta-DBSs (n)	Associated genes (n)
Individual analysis	GSE94023	MCF7	E2	10	38,963	13,982
	GSE99626				15,372	9,013
	GSE67295				29,145	11,531
	GSE115607				21,939	11,020
	GSE80367	T47D			4,877	3,081
	GSE23893	T47D		100	1,882	1,320
	GSE23893	MCF7			11,155	6,272
	GSE54855				10,343	5,710
Meta-analysis	GSE59530				29,915	10,817
	GSE94023	MCF7		10	23,308	10,503
	GSE99626					
	GSE67295					
	GSE115607					
	GSE23893			100	7,796	4,591
	GSE54855					
	GSE59530					

DBSs differentially bound sites, Meta-DBSs meta-differentially bound sites, nM Nanomolar

ChIP-seq data were meta-analyzed, and the meta-DBSs ( $-\log_{10} (q\text{-value}) > 2$ ) were identified (see Additional file 2: Tables S12 and S13). The details of DBSs, meta-DBSs, and associated genes are described in Table 2.

#### Shared significant TFBSs between meta-DBSs and DBSs

Meta-DBSs were compared with DBSs obtained from individual datasets for more precision. Based on our findings, 617 genes were common among meta-DBSs- and DBSs- associated genes, 35 of which were TFs (Fig. 4). Remarkably, nine of those TFs were also identified among the top 50 TFs obtained from ChEA3 in Section 3.5.1 (Fig. 6). There were 282 genes associated with peaks identified through *meta-analysis* of MCF7 cell lines treated with 10 nM E2 but not in initial individual datasets. Six of them were *POU5F1B*, *ZNF662*, *ZNF442*, *KIN*, *ZNF410*, and *SGSM2* TFs (Fig. 4A). Moreover, most of the top 50 TFs were also identified among the *meta-analysis* results and some initial individual datasets (Figs. 4 and 6). *PCGF2*, *HNF1B*, and *ZBED6* TFs predicted by ChEA3 were not identified through *meta-analysis* or initial individual datasets.

#### Genomic occupancy of ER binding sites

The genomic locations of 7,308 ER-meta-DBSs correlated with 617 common genes were annotated using the ChIPseeker. According to the genome-wide annotation of TFBSs, it was found that there were 1,534 binding sites located within 10 kb of a transcription start site

(TSS) (Fig. 5A). 1,531 of these sites were situated in the proximal to the TSSs (promoter region). Also, peaks were 3,327 in introns, 2,104 in intergenic, 218 in exons, 13 in 5' untranslated regions (UTRs), three in TSSs (downstream region), and 112 in 3' UTRs (Fig. 5B and C). Because some of the annotations overlap, the complete annotations and their overlaps are illustrated in Fig. 5D. These results indicated that most ER binding sites are located in introns, followed by intergenic regions, promoter-TSSs, and exons to a lesser degree in 3' UTRs, TSSs, and 5' UTRs.

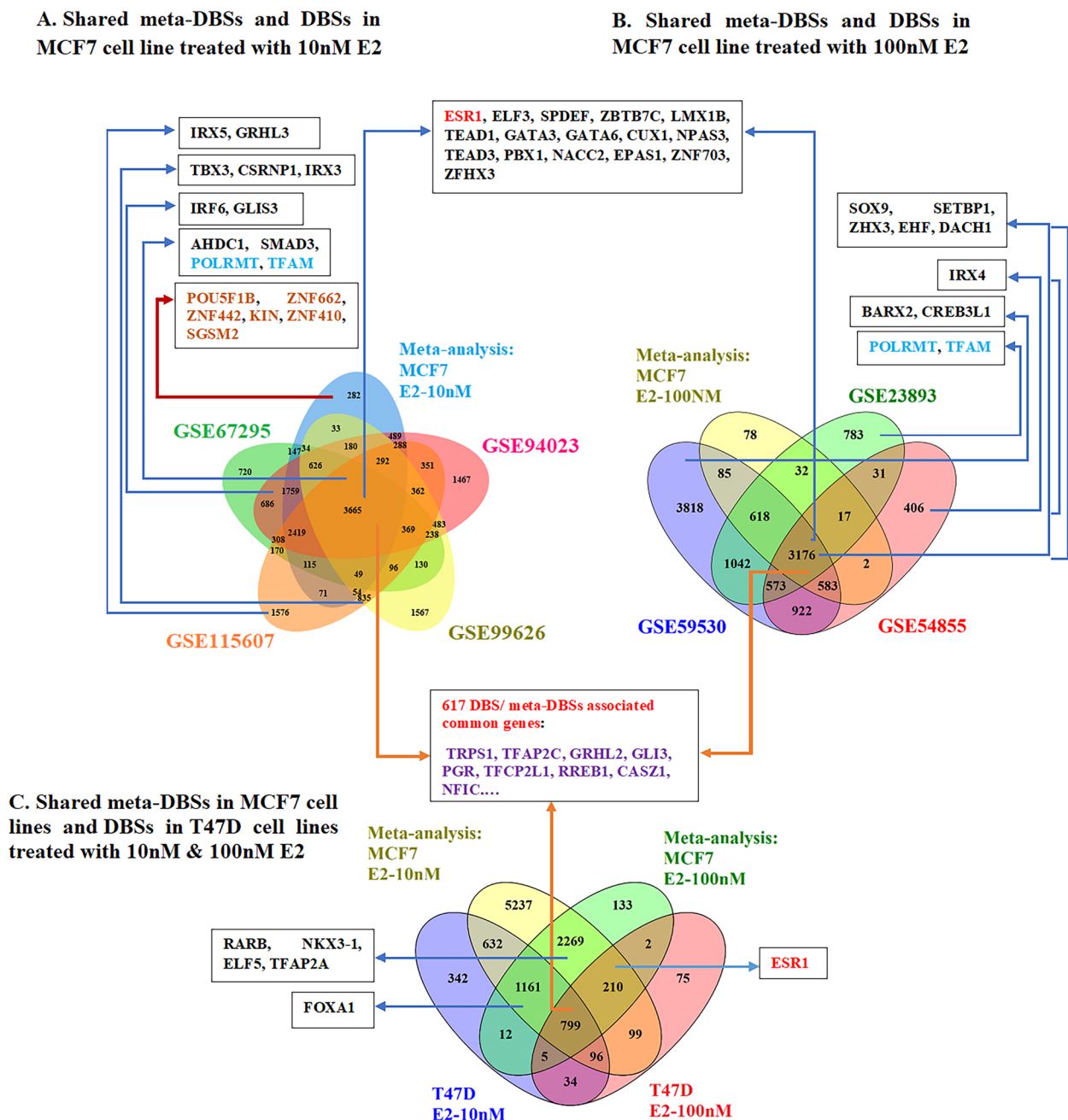
#### Enrichment analysis

##### ChIP enrichment analysis

The 617 meta-DBSs-associated common genes were enriched using the ChEA3 database to predict significant TFs that regulate gene expression in breast cancer. The results of six libraries and two overall results are presented in Additional file 2: Tables S14-S21. In accordance with the integrated\_meanRank (see Additional file 2: Table S21), 1632 TFs were ranked, and the top 50 TFs are shown in Fig. 6. Remarkably, *TRPS1*, *FOXA1*, *TFAP2C*, *GLIS3*, *ELF3*, and *ESR1* TFs have the highest rank score. Also, *PCGF2*, *HNF1B*, and *ZBED6* TFs were predicted as potential key regulators.

##### Peak set functional enrichment analysis

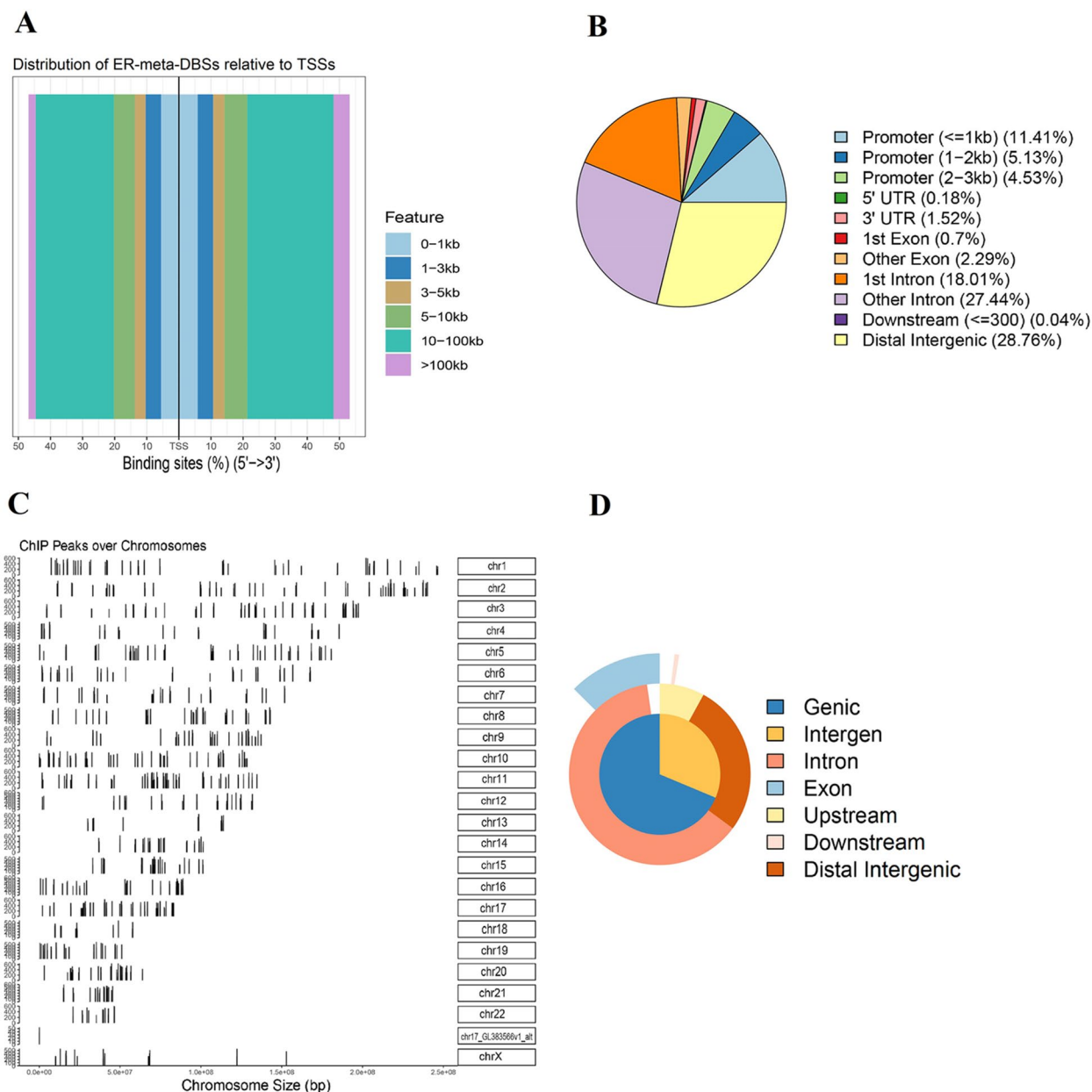
Peak set functional enrichment analysis, including GO and KEGG pathways for 7,308 meta-DBSs (correlated



**Fig. 4** Shared meta-DBSs between meta-analysis and individual studies based on TFBSs associated genes in MCF7 and T47D cell lines treated with E2. Among the peaks associated genes, only TFs are displayed. Nine of these TFs, which are also among the top 50 TFs obtained from ChEA3, are shown in purple. The six TFs only identified through meta-analysis are shown in orange. The marked blue TFs are mitochondrial TFs. TFs Transcription factors, DBSs Differentially bound sites, Meta-DBSs Meta-differentially bound sites

with common 617 genes) was performed. By enriching meta-DBSs and considering adjusted  $P$ -value  $\leq 0.05$ , 94 BPs, 3 CCs, and 12 MFs for GO terms and 9 KEGG pathways were detected (see Additional file 2: Tables S22 and S23). GO terms regarding BPs were annotated in cornification (GO:0070268), tissue development (GO:0009888),

regulation of anatomical structure morphogenesis (GO:0022603), anterior/posterior pattern specification (GO:0009952), phosphatidylcholine metabolic process (GO:0046470), negative regulation of transcription by RNA polymerase II (GO:0000122), response to metal ion (GO:0010038), negative regulation of RNA biosynthetic



**Fig. 5** Genome-wide annotation of 7,308 meta-DBSs correlated with 617 common genes and response elements of ER between MCF7 and T47D cell lines treated with E2. **A** Distribution of ER-meta-DBSs relative to the nearest TSS across the human genome. **B** Pie plot of the ER-meta-DBSs percentages according to peak location across different genomic regions of the human genome. **C** Visualization of ER-meta-DBSs obtained from the UCSC genome browser (version hg38). **D** Venn pie of annotations and their overlap. TSS transcription start site

process (GO:1902679), embryonic organ morphogenesis (GO:0048562), and response to inorganic substance (GO:0010035).

Also, enriched CCs terms were linked to the cortical actin cytoskeleton (GO:0030864), cortical cytoskeleton (GO:0030863), and cytoskeleton (GO:0005856). Significant GO terms were found for MFs with regulatory region nucleic acid binding (GO:0001067), glucose

binding (GO:0005536), lipid binding (GO:0008289), RNA polymerase II regulatory region DNA binding (GO:0001012), activating transcription factor binding (GO:0033613), DNA-binding transcription factor binding (GO:0140297), and transcription coregulator activity (GO:0003712) (Fig. 7A).

The nine enriched KEGG pathways included the Estrogen signaling pathway (hsa04915), Rap1 signaling





**Fig. 6** ChIP Enrichment Analysis (ChEA) for 617 meta-DBS-associated common genes. The bar graph integrated *integrated\_meanRank* for the top 50 TFs using ChEA3 in MCF7 and T47D cell lines treated with E2. Each color represents a library, and each bar's length indicates the weight of that TF in each library. *Tfs* transcription Factors

pathway (hsa04015), Tight junction (hsa04530), Neurotrophin signaling pathway (hsa04722), Fluid shear stress and atherosclerosis (hsa05418), MAPK signaling pathway (hsa04010), Thyroid cancer (hsa05216), Bladder cancer (hsa05219), and Pathways in cancer (hsa05200) (Fig. 7B). Moreover, the top 50 TFs obtained from ChEA3 were examined. The TFs that enriched the Top 20 GO terms and nine KEGG pathways with the number of peaks in each route is shown in Fig. 7.

A general classification of pathways obtained from the enrichment of meta-DBSs indicated that most of those pathways play a role in cancer, environmental information processing, and organismal systems (Fig. 8).

**Evaluation of ChIP-seq meta-analysis results along with RNA-seq meta-analysis results**

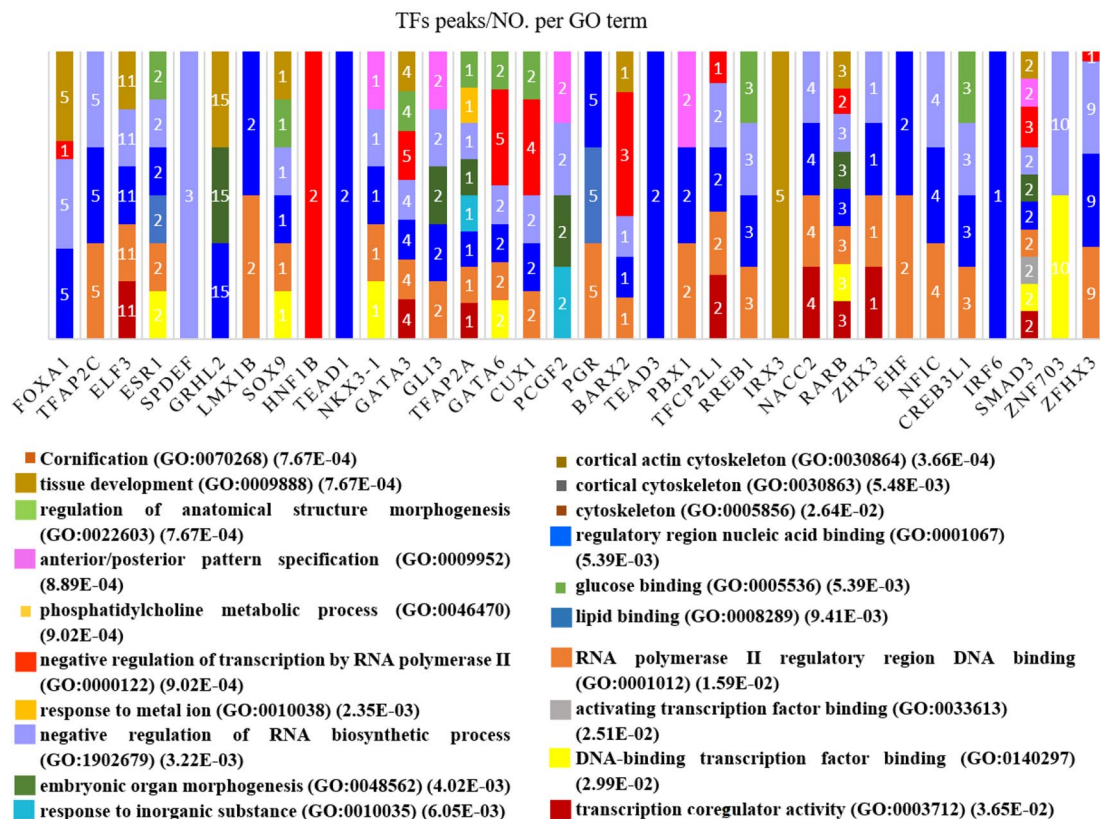
Recently we performed a *meta-analysis* of RNA-seq data on both MCF7 and T47D cell lines treated with 10 nM E2 for 8 h [32]. We investigated the expression of the top 50 TFs predicted by ChEA3 (Table 3). Comparing *meta-analysis* results of ChIP-seq and RNA-seq data showed that many TFs were up-regulated in RNA-seq *meta-analysis* or individual datasets.

**Discussion**

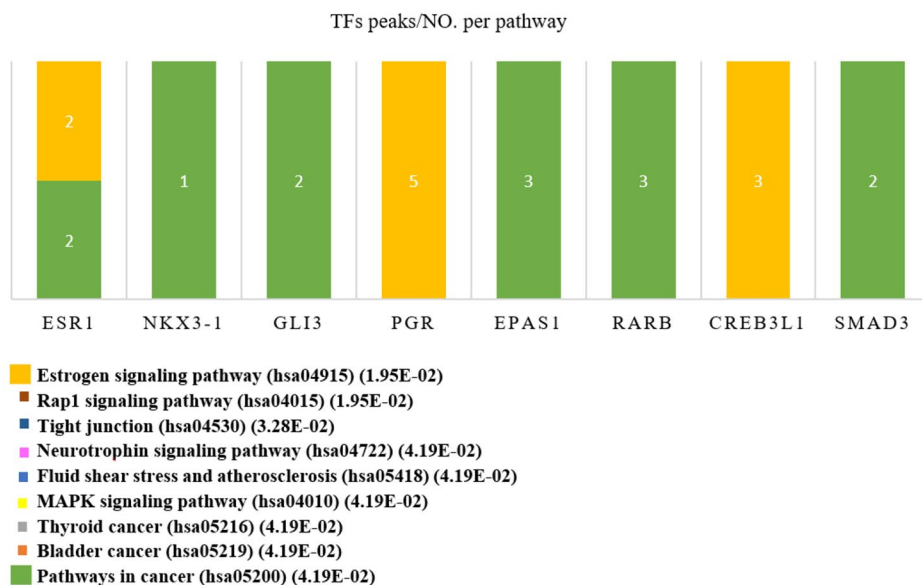
Breast cancer is the most incident cancer among women worldwide. The largest group of patients with BC is ER<sup>+</sup>. The estrogen receptor plays the role of a transcription

factor when exposed to estrogen and triggers cellular proliferation and differentiation [2, 3]. Hence, evaluating the expression of TFs directly affected by ER can help identify and understand ER action's mechanism in cancer progression. Remarkably, combining more samples and performing *meta-analysis* will produce more stringent results than single experiments [9]. To the best of our knowledge in the current study, for the first time, we applied the workflow for *meta-analysis* of ChIP-seq data based on TFBSs in the MCF7 cell lines treated with 10 nM and 100 nM E2. First, samples were integrated, and known and unknown batch effects were removed. Next, the *meta-analysis* approach was conducted on TFBSs scores, and meta-DBSs were obtained. Our results showed significant peaks (specifically for TFs affected by ER) identified through the *meta-analysis* but not found in individual studies. For the strict selection of meta-DBSs, MCF7 cell lines *meta-analysis* results were compared with the results of T47D cell lines and initial individual studies, and common genes and the associated meta-DBSs were determined. With the enrichment of meta-DBSs correlated with common genes, several pathways with presumably clinical importance in BC were identified. Meta-DBSs-associated common genes were also enriched, and TFs were identified that have been predicted as potential key regulators in the progression of BC. In addition, we compared the *meta-analysis* results of both ChIP-seq and RNA-seq data.

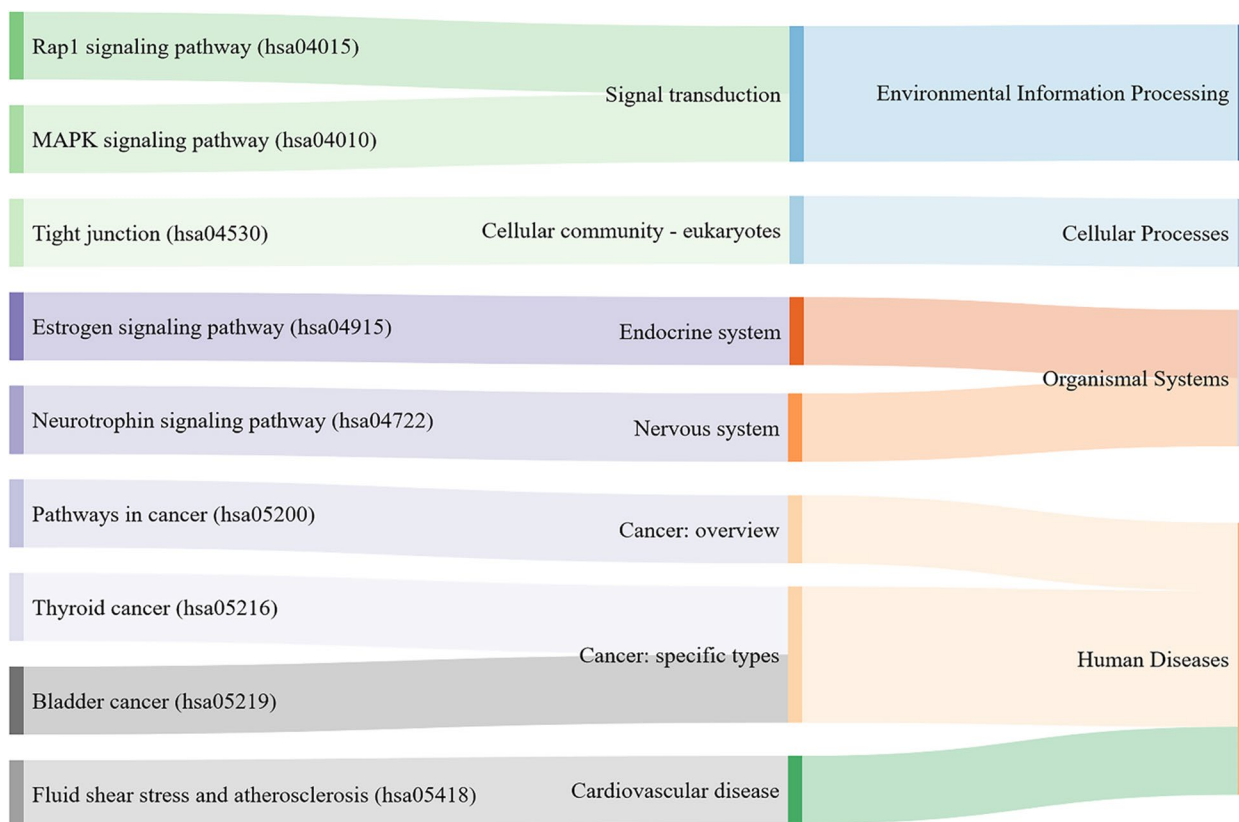
### A. Gene Ontology



### B. KEGG pathways



**Fig. 7** Peak set functional enrichment analysis (PSFEA) for 7,308 meta-DBSs, including GO and KEGG pathways. **(A)** The chart of the top 20 GO terms and **(B)** nine KEGG pathways was obtained from 7,308 meta-DBSs in MCF7 and T47D cell lines treated with E2. Also, TFs that enriched GO terms and KEGG pathways are shown along with the number of peaks. *TFs* Transcription factors, *DBSs* Differentially bound sites, *Meta-DBSs* Meta-differentially bound sites



**Fig. 8** Sankey diagram for categories of KEGG pathways

Some of the differentially expressed genes (DEGs) identification methods in the RNA-seq context, such as DESeq2 [53] and edgeR [54], are also exploited to discover DBSs. However, the NOIseq package [26, 28] performs relatively better for obtaining DEGs [28]. Moreover, packages were developed for *meta-analysis* of RNA-seq data based on those methods [28, 55, 56]. To find meta-DBSs, we utilized the metaSeq package [28], in which NOIseq has been used. Before performing the *meta-analysis*, we faced two problems: (1) none of the mentioned methods removed unknown batch effects, and (2) the number of peaks and their associated genomic start and end intervals were different among samples, and batch effects could not be removed. Therefore, sample preparation was required. To this end, we obtained the same number, regions, and scores of TFBSs using the DiffBind package.

Herein, *POU5F1B*, *ZNF662*, *ZNF442*, *KIN*, *ZNF410*, and *SGSM2* TFs were found in the *meta-analysis* of MCF7 cell lines treated with 10 nM E2 but not in individual studies. *POU5F1B*, a processed pseudogene highly homologous to *OCT4*, was recently shown to be transcribed in ER<sup>+</sup> BC [57]. *POU5F1B* is specifically expressed in mammalian totipotent embryonic stem and germ cells and has a crucial role in regulating and

maintaining pluripotency and self-renewal [57]. The Zinc finger protein family often functions in transcriptional regulation through sequence-specific DNA binding [58]. *ZNF662*, *ZNF442*, and *ZNF410* were involved in the regulation of transcription by RNA polymerase II [59]. It has shown that low expression of G protein-coupled estrogen receptor 1 (*GPER1*) is significantly associated with adverse survival of BC patients [60]. *ZNF662* was identified as one of the unique factors related to *GPER*-DNA binding [60]. *ZNF442* plays a role in the strategy adopted by ER<sup>+</sup> BC and triple-negative breast cancer (TNBC) cell lines for maintaining zinc homeostasis [61]. *ZNF410* uniquely activates *CHD4*, one of the catalytic components of the Nucleosome Remodeling and Deacetylase (NuRD) complex [62]. NuRD has been shown to have opposing effects on cancer. For instance, promoting and inhibiting tumor growth and metastasis in different tissues such as ER<sup>+</sup> BC [63, 64]. Recently, it has been shown that the Knockdown of DNA/RNA-binding protein KIN17 (*KIN*) promotes apoptosis of TNBC [65]. Small G protein signaling modulator 2 (*SGSM2*) is involved in ER<sup>+</sup> BC metastasis by enhancing migrator cell adhesion via interaction with E-cadherin [66].

**Table 3** Differential express of top 50 TFs using ChEA3 for 617 meta-DBSs-associated common genes [32–52]

TFs	DEGs in MCF7 or/ & T47D cell lines	Meta-DEGs in MCF7 cell lines	Meta-DEGs in T47D cell lines	DEGs in TNBC cell lines	Experiment	TFs	DEGs in MCF7 or/ & T47D cell lines	Meta-DEGs in MCF7 cell lines	Meta-DEGs in T47D cell lines	DEGs in TNBC cell lines	Experiment
TRPS1	-	-	up	-	Piryaei, ... et al.[32]	NPAS3	down	down	-	-	Piryaei, ... et al.[32]
FOXA1	up	-	-	up	Kumar, ... et al.[33] Mohammed, ... et al.[34]	PGR	up	up	up	-	Piryaei, ... et al.[32]
TFAP2C	-	up	up	-	Piryaei, ... et al.[32]	BARX2	up	up	-	-	Piryaei, ... et al.[32]
GLIS3	-	-	-	up	Rami, ... et al.[35]	TEAD3	up	-	-	-	Ma, ... et al.[36]
ELF3	-	down	-	-	Luk, ... et al.[37]	PBX1	-	-	down	-	Piryaei, ... et al.[32]
ESR1	-	-	up	-	Piryaei, ... et al.[32]	CSRNP1	-	down	-	-	Piryaei, ... et al.[32]
ZBED6	-	-	-	up	Cai, ... et al.[49]	TFCP2L1	up	up	-	-	Piryaei, ... et al.[32]
SPDEF	down	down	-	-	Piryaei, ... et al.[32]	RREB1	-	-	-	up	Deng, ... et al.[38]
GRHL2	-	up	up	-	Piryaei, ... et al.[32]	IRX3	up	up	up	-	Piryaei, ... et al.[32]
ZBTB7C	up	up	-	-	Piryaei, ... et al.[32]	NACC2	down	-	-	-	Piryaei, ... et al.[32]
LMX1B	-	up	up	-	Piryaei, ... et al.[32]	SETBP1	-	-	-	up	Chen, ... et al.[39]
TBX3	down	down	-	-	Piryaei, ... et al.[32]	EPAS1	-	-	down	-	Piryaei, ... et al.[32]
AHDC1	up	-	-	-	Honkela, ... et al.[40]	RARB	up	-	-	up	Lubecka, ... et al.[41]
IRX5	-	-	-	up	Singh, ... et al.[42]	ZHX3	-	-	-	up	Hafez, ... et al.[43]
SOX9	down	-	-	-	Piryaei, ... et al.[32]	CASZ1	down	-	-	-	Li, ... et al.[44]
HNF1B	-	-	-	up	Pilato, ... et al.[45]	EHF	down	-	-	-	Luk, ... et al.[37]
TEAD1	up	-	-	-	Ma, ... et al.[36]	DACH1	up	-	-	up	Zhao, ... et al.[46]
NKX3-1	up	-	-	-	Asch-Kendrick, ... et al.[47]	NFIC	-	up	up	-	Piryaei, ... et al.[32]
GATA3	-	down	-	-	Piryaei, ... et al.[32]	CREB3L1	-	-	-	up	Denard, ... et al.[48]
GLI3	-	-	up	-	Piryaei, ... et al.[32]	IRX4	-	-	-	up	Corrêa, ... et al.[49]
TFAP2A	up	-	-	up	Xu, ... et al.[50]	IRF6	down	down	-	-	Piryaei, ... et al.[32]
ELF5	down	down	-	-	Piryaei, ... et al.[32]	SMAD3	up	-	-	-	Deng, ... et al.[38]
GATA6	-	-	-	up	Song, ... et al.[51]	GRHL3	down	down	down	-	Piryaei, ... et al.[32]
CUX1	-	up	-	-	Piryaei, ... et al.[32]	ZNF703	up	up	up	-	Piryaei, ... et al.[32]
PCGF2	-	-	-	up	Khan, ... et al.[52]	ZFH3	-	down	-	-	Piryaei, ... et al.[32]

TFs Transcription factors, Meta-DEGs Meta-differentially expressed genes, TNBC Triple-negative breast cancer

TFs Transcription factors, Meta-DEGs Meta-differentially expressed genes, TNBC Triple-negative breast cancer

With enrichment analysis of meta-DBSs-associated common genes by ChEA3, TFs including *PCGF2*, *HNF1B*, and *ZBED6* were predicted as regulatory factors involved in BC progression. Based on our findings, no peaks associated with those TFs were found in individual studies and *meta-analysis* results. Maybe it is because all TFs do not always regulate their own expression. Polycomb group proteins (PcG) play a critical role in cancer development, proliferation, senescence, and carcinogenesis [67]. *PCGF2* serves as a tumor suppressor in BC, gastric cancer, and colon cancer, probably for the negative regulation of Akt activation [68]. Recent studies

have shown that methylation of homeobox genes, such as the *HNF1B*, plays a critical role in BC's insurgence or progression [69]. Zinc Finger BED-Type Containing 6 (*ZBED6*) has been shown to repress insulin-like growth factor 2 (*IGF2*) transcription [59]. The role of *ZBED6* has only been reported in TNBC but is not well understood in ER<sup>+</sup> BC so far [70].

Most TFs indicated in Fig. 4 were shared among results of individual studies, *meta-analysis*, and the top 50 TFs obtained from ChEA3 (Fig. 6). The *ESR1* and *PGR*, which play an essential role in the development of BC, were ranked as top-ranking TFs (Fig. 6).



Because ChEA3 ranks TFs only based on a gene list without any assumptions, it confirms the correctness of our results. It verifies that removing batch effects and performing a correct *meta-analysis* can produce more exact results. *POLRMT* and *TFAM* are human mitochondrial TFs [71]. As an essential mitochondrial DNA (mtDNA)-binding protein, *TFAM* functions in genome maintenance such as determining the abundance of the mitochondrial genome. Emerging evidence indicates that altered *TFAM* levels or mtDNA copy numbers may impact mitochondrial homeostasis in Alzheimer's and other neurodegenerative diseases [71]. Several studies suggested that *TFAM* TFs are potential biomarkers for the prognosis of BC [72]. *POLRMT* is a mitochondrial RNA polymerase that requires *TFAM* for mitochondrial transcription initiation [73]. Targeting both those factors is a promising strategy for sensitizing BC cells to chemotherapy [72, 74]. Although most of the datasets analyzed in the present study have used the chip-anti-ER (Santa Cruz HC-20, cat# SC-543), the tagging approach could alter the binding preferences of the protein, and therefore the experimental results.

Interestingly, the significant GO terms and KEGG pathways obtained from PSFEA in the present study were enriched by many predicted TFs by ChEA3 (Figs. 6 and 7). Therefore, *meta-analysis* led to the recognition of TFs involved in the development of BC. Cornification (GO:0070268) [75], anterior/posterior pattern specification (GO:0009952) [76], embryonic organ morphogenesis (GO:0048562) [77], response to inorganic substance (GO:0010035) [78], and negative regulation of transcription by RNA polymerase II (GO:0000122) [79] have been previously reported in ER<sup>+</sup> BC. Tissue development (GO:0009888) [80] was enriched in BC. Cortical actin cytoskeleton (GO:0030864) [81], and cortical cytoskeleton (GO:0030863) [81] in pancreatic cancer, and cytoskeleton (GO:0005856) [82, 83] in ovarian cancer were enriched but not reported in ER<sup>+</sup> BC previously. Regulation of anatomical structure morphogenesis (GO:0022603) [84] in colorectal cancer and phosphatidylcholine metabolic process (GO:0046470) [85] in hepatocellular carcinoma were enriched. Response to metal ion (GO:0010038) [86] in chronic myeloid leukemia cells, and negative regulation of RNA biosynthetic process (GO:1902679) [87] in Alzheimer's Disease, were enriched but not reported in BC before. Regulatory region nucleic acid binding (GO:0001067) [88], lipid binding (GO:0008289) [89], RNA polymerase II regulatory region DNA binding (GO:0001012) [90], activating transcription factor binding (GO:0033613) [91], DNA-binding transcription factor binding (GO:0140297) [92], transcription coregulator activity (GO:0003712) [93] have been previously reported in ER<sup>+</sup> BC. Glucose binding

(GO:0005536) [94] was enriched in multiple myeloma but not reported in BC before.

The nine enriched KEGG pathways include Pathways in cancer (hsa05200) [95], Estrogen signaling pathway (hsa04915) [95], Rap1 signaling pathway (hsa04015) [96], Neurotrophin signaling pathway (hsa04722) [97], and Fluid shear stress and atherosclerosis (hsa05418) [98] have been previously reported in ER<sup>+</sup> BC. MAPK signaling pathway (hsa04010) [99] was enriched in MCF7 cell lines of the resistance to treatment. Tight junction (hsa04530) [100], and Bladder cancer (hsa05219) [101, 102] have already been reported tumor BC. A study has shown that Thyroid cancer (hsa05216) [103] is related to metabolic pathways which were enriched in TNBC.

According to the cases mentioned, GO terms and pathways previously approved concerning BC were enriched (Fig. 7). Moreover, many GO terms and pathways identified were associated with gastrointestinal cancers, including colorectal and hepatocellular carcinoma. In particular, our findings showed that many TFBSs associated with mitochondrial genes are directly affected by ER, including *POLRMT* and *TFAM* TFs. Especially *TFAM* increased expression has been reported in a *meta-analysis* of MCF7 cell lines treated with E2 [32]. This may clarify why the metabolism pathways were identified along with the estrogen signaling pathway in BC (Fig. 7 and see Additional file 2: Table S22). Although some relationships are found between BC and hepatocellular carcinoma, more studies are needed on these TFs and their associated pathways.

The comparison of ChIP-seq and RNA-seq data *meta-analysis* results showed that many TFs predicted by ChEA3 were up-regulated in RNA-seq *meta-analysis* results. There can be several reasons why some TFs are down-regulated. *FOXAI* is a determinant of drug resistance in breast cancer cells, and it has different functions in response to treatment in ER<sup>+</sup> and TNBC cell lines [33] (Table 3). Investigation of primary regulatory regions has shown that there is plasticity in ER binding, with distinct ER binding profiles associated with clinical outcome. These differential ER binding profiles appear to be mediated by *FOXAI* [104].

*SPDEF* also is with both oncogenic and tumor-suppressor functions in BC [105]. Therefore, according to its role in the cell, it can have a different expression. Furthermore, the treatment period of RNA-seq data can be effective in the gene expressions, so it is possible that at one time, the expression of the gene is increased, and at another time, it is down-regulated.

Our study was constrained by the number of datasets with the same conditions. Although the results were remarkable and substantial, they could be more precise and reliable if more data were combined. In addition,



despite generating large volumes of datasets related to BC in databases, the data with the same conditions is negligible. Therefore, developing datasets with more uniform and focused conditions is could help to detect more precise *meta-analysis* results.

## Conclusion

In the current study, we applied a workflow for integrating and performing *meta-analysis* of several ChIP-seq studies based on TFBSs scores in ER<sup>+</sup> BC. The same number, regions, and scores of TFBSs were obtained for all samples. Then, the known and unknown batch effects were removed, and the *meta-analysis* was performed to obtain meta-DBSs. Some TFs were significantly identified in the *meta-analysis* but not in individual studies. Also, with meta-DBSs enrichment, many GO terms and pathways were identified that were not previously reported in BC. Finally, by enriching the meta-DBSs-associated genes, some TFs were prioritized and predicted as potential regulators. Those TFs were not found in the results of the *meta-analysis* and individual studies. Moreover, the results of the ChIP-seq *meta-analysis* were confirmed by comparing them with the *meta-analysis* of the RNA-seq data. The results showed that many TFs predicted by ChEA3 were up-regulated in RNA-seq *meta-analysis* or individual datasets results. As a suggestion, this workflow can be applied to other types of BC, such as HER2<sup>+</sup> and other diseases, provided that there are sufficient datasets with the same conditions. It can also be performed on other TFs involved in BC that cooperate with ER as a cofactor, revealing new insights into ER mechanism of action.

## Abbreviations

BC	Breast cancer
DEGs	Differentially expressed genes
ER <sup>+</sup>	Estrogen receptor positive
HER2 <sup>+</sup>	Human epidermal growth factor receptor 2
GSEA	Gene set enrichment analysis
GO	Gene ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
Meta-DEGs	Meta-differentially expressed genes
TNBC	Triple negative breast cancer
TFBSs	Transcription factor binding sites
DBSs	Differentially bound sites
Meta-DBSs	Meta-differentially bound sites
PSFEA	Peak set functional enrichment analysis
ChEA	ChIP Enrichment Analysis
TSS	Transcription start site

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12920-023-01655-z>.

**Additional file 1: Figure S1.** (A) The Heat map Correlation matrix and (B) PCA plot for the GSE94023 study. MCF7 cell line was treated with E2 for different times, including 0 (treated with vehicle as control), 5, 10, 20,

40, 80, 160, 320, 640, and 1280 minutes. According to the matrix, after 40 minutes, E2-treated samples showed a more significant difference compared to the control sample. **Figure S2.** The Heat map Correlation matrix for GSE94023, GSE99626, GSE67295, and GSE115607 studies. The default binding affinity matrix was obtained from the DiffBind package. The datasets with MCF7 cell lines that were treated with 10nM E2 for 40 or 45 minutes were shown (A) before and (B) after normalization and unknown batch effect correction. Batch effect removal was performed using the ARSvNseq package in R. **Figure S3.** The PCA plot for GSE94023, GSE99626, GSE67295, and GSE115607 studies. The default binding affinity matrix was obtained from the DiffBind package. The datasets with MCF7 cell lines that were treated with 10nM E2 for 40 or 45 minutes were shown (A) before and (B) after normalization and unknown batch effect correction. **Figure S4.** The Heat map Correlation matrix for GSE23893, GSE54855, and GSE59530 studies. The default binding affinity matrix was obtained from the DiffBind package. The datasets with MCF7 cell lines that were treated with 100nM E2 for 40 or 45 minutes were shown (A) before and (B) after normalization and unknown batch effect correction. Batch effect removal was performed using the ARSvNseq package in R. **Figure S5.** The PCA plot for GSE23893, GSE54855, and GSE59530 studies. The default binding affinity matrix was obtained from the DiffBind package. The datasets with MCF7 cell lines that were treated with 100nM E2 for 40 or 45 minutes were shown (A) before and (B) after normalization and unknown batch effect correction.

**Additional file 2: Table S1.** Differentially bound sites (DBSs) obtained from MCF7 cell line treated with 10nM E2 for 45 minutes in GSE94023 study. **Table S2.** Differentially bound sites (DBSs) obtained from MCF7 cell line treated with 10nM E2 for 45 minutes in GSE99626 study. **Table S3.** Differentially bound sites (DBSs) obtained from MCF7 cell line treated with 10nM E2 for 45 minutes in GSE67295 study. **Table S4.** Differentially bound sites (DBSs) obtained from MCF7 cell line treated with 10nM E2 for 45 minutes in GSE115607 study. **Table S5.** Differentially bound sites (DBSs) obtained from T47D cell line treated with 10nM E2 for 45 minutes in GSE80367 study. **Table S6.** Differentially bound sites (DBSs) obtained from T47D cell line treated with 100nM E2 for 45 minutes in GSE23893 study. **Table S7.** Differentially bound sites (DBSs) obtained from MCF7 cell line treated with 100nM E2 for 45 minutes in GSE23893 study. **Table S8.** Differentially bound sites (DBSs) obtained from MCF7 cell line treated with 100nM E2 for 45 minutes in GSE54855 study. **Table S9.** Differentially bound sites (DBSs) obtained from MCF7 cell line treated with 100nM E2 for 45 minutes in GSE59530 study. **Table S10.** Default binding affinity matrix of 6 samples by the 63,612 sites that overlap in at least two of the samples using DiffBind in (GSE94023, GSE99626, GSE67295, & GSE115607) MCF7 cell line treated with 10nM E2 for 45 minutes. **Table S11.** Default binding affinity matrix of 6 samples by the 23,517 sites that overlap in at least two of the samples using DiffBind in (GSE23893, GSE54855, & GSE59530) MCF7 cell line treated with 100nM E2 for 45 minutes. **Table S12.** Meta-differentially bound sites (meta-DBSs) obtained from a meta-analysis on (GSE94023, GSE99626, GSE67295, & GSE115607) MCF7 cell line treated with 10nM E2 for 45 minutes. **Table S13.** Meta-differentially bound sites (meta-DBSs) obtained from a meta-analysis on (GSE23893, GSE54855, & GSE59530) MCF7 cell line treated with 100nM E2 for 45 minutes. **Table S14.** literature\_ChIP-seq. **Table S15.** Enrichr. **Table S16.** ARCHS4—Coexpression. **Table S17.** ENCODE—ChIP-seq. **Table S18.** ReMap—ChIP-seq. **Table S19.** GTEx—Coexpression. **Table S20.** Integrated\_topRank. **Table S21.** Integrated\_meanRank. **Table S22.** Gene Ontology (GO) for 7,308 meta-DBSs related to 617 common genes among MCF7 & T47D cell lines using Cistrome-GO. **Table S23.** KEGG pathways analysis for 7,308 meta-DBSs related to 617 common genes among MCF7 & T47D cell lines using Cistrome-GO. **Table S24.** Differentially expressed genes (DEGs) identified from GRO-seq data in the MCF7 cell line treated with 100nM E2 for 40 minutes in the GSE27463 study.

## Acknowledgements

Not applicable.

## Authors' contributions

Z.P. searched and collected preliminary studies from databases. Z.P. and Z.S. investigated the datasets. Z.P. downloaded and performed initial analyzes on

ChIP-seq raw data. Z.P. and K.K. designed the workflow for performing the ChIP-seq data meta-analysis. Z.P. performed statistical and computational analysis and the interpretation of results and wrote the article. Z.P., Z.S., and K.K. contributed to editing the article. E.E. and M.E. proposed the initial idea of this research.

### Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### Availability of data and materials

The datasets used in the current study including GSE94023, GSE99626, GSE67295, GSE115607, GSE80367, GSE23893, GSE54855, and GSE59530 are publicly available and were collected from SRA-NCBI (<https://www.ncbi.nlm.nih.gov/>) database. They were downloaded from ENA-EBI (<https://www.ebi.ac.uk/ena/browser/>) database.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>Department of Bioinformatics, Kish International Campus University of Tehran, Kish, Iran. <sup>2</sup>Laboratory of Complex Biological Systems and Bioinformatics (CBB), Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran. <sup>3</sup>Hematology-Oncology and Stem Cell Transplantation Research Center, Tehran University of Medical Sciences, Tehran, Iran. <sup>4</sup>Genomics Research Platform, School of Agriculture, Biomedicine and Environment, La Trobe University, Melbourne, VIC, Australia. <sup>5</sup>School of Animal and Veterinary Sciences, The University of Adelaide, Adelaide, South Australia, Australia.

Received: 24 February 2023 Accepted: 4 September 2023

Published online: 15 September 2023

### References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209–49. <https://doi.org/10.3322/caac.21660>.
- Brosens JJ, Tullet J, Varshochi R, Lam EWF. Steroid receptor action. *Best Pract Res Clin Obstet Gynaecol.* 2004;18(2):265–83. <https://doi.org/10.1016/j.bpobgyn.2004.01.006>.
- Fuller PJ. The steroid receptor superfamily: mechanisms of diversity. *FASEB J.* 1991;5(15):3092–9. <https://doi.org/10.1111/febs.12658>.
- Liao XH, Lu DL, Wang N, Liu LY, Wang Y, Li YQ, et al. Estrogen receptor  $\alpha$  mediates proliferation of breast cancer MCF-7 cells via a p21/PCNA/E2F 1-dependent pathway. *FEBS J.* 2014;281(3):927–42. <https://doi.org/10.1111/febs.12658>.
- Bono H, Hirota K. Meta-analysis of hypoxic transcriptomes from public databases. *Biomedicines.* 2020;8(1):10. <https://doi.org/10.3390/biomedicines8010010>.
- Puyang X, Furman C, Zheng GZ, Wu ZJ, Banka D, Aithal K, et al. Discovery of selective estrogen receptor covalent antagonists for the treatment of ER $\alpha$ WT and ER $\alpha$ MUT breast cancer. *Cancer Discov.* 2018;8(9):1176–93. <https://doi.org/10.1158/2159-8290.CD-17-1229>.
- Dzida T, Iqbal M, Charapitsa I, Reid G, Stunnenberg H, Matarese F, et al. Predicting stimulation-dependent enhancer-promoter interactions from ChIP-Seq time course data. *PeerJ.* 2017;5:e3742. <https://doi.org/10.7717/peerj.3742>.
- Stender JD, Nwachukwu JC, Kastrati I, Kim Y, Strid T, Yakir M, et al. Structural and molecular mechanisms of cytokine-mediated endocrine resistance in human breast cancer cells. *Molecular Cell.* 2017;65(6):1122–35. e5. <https://doi.org/10.1016/j.molcel.2017.02.008>.
- Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*: Lippincott Williams & Wilkins; 2008. <https://doi.org/10.1016/j.annemergmed.2008.06.461>.
- Kolmykov SK, Kondrakhin YV, Sharipov RN, Yevsh IS, Ryabova AS, Kolpakov FA, editors. *Meta-analysis of ChIP-seq Datasets through the Rank Aggregation Approach*. 2020 Cognitive Sciences, Genomics and Bioinformatics (CSGB). IEEE; 2020. <https://doi.org/10.1109/CSGB51356.2020.9214614>.
- Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 2021;49(D1):D10. <https://doi.org/10.1093/nar/gkaa892>.
- Cochrane G, Alako B, Amid C, Bower L, Cerdeño-Tárraga A, Cleland I, et al. Facing growth in the European nucleotide archive. *Nucleic Acids Res.* 2012;41(D1):D30–5. <https://doi.org/10.1093/nar/gks1175>.
- Jalili V, Afgan E, Gu Q, Clements D, Blankenberg D, Goecks J, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Res.* 2020;48(W1):W395–402. <https://doi.org/10.1093/nar/gkaa434>.
- Andrews S. FastQC: a quality control tool for high throughput sequence data. 0.11.5. Babraham Bioinformatics. 2016. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
- Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015;12(4):357–60. <https://doi.org/10.1038/nmeth.3317>.
- Singhal H, Greene ME, Zarnke AL, Laine M, Al Abosy R, Chang Y-F, et al. Progesterone receptor isoforms, agonists and antagonists differentially reprogram estrogen signaling. *Oncotarget.* 2018;9(4):4282. <https://doi.org/10.18632/oncotarget.21378>.
- Singhal H, Greene ME, Tarulli G, Zarnke AL, Bourgo RJ, Laine M, et al. Genomic agonism and phenotypic antagonism between estrogen and progesterone receptors in breast cancer. *Sci Adv.* 2016;2(6):e1501924. <https://doi.org/10.1126/sciadv.1501924>.
- Kong SL, Li G, Loh SL, Sung WK, Liu ET. Cellular reprogramming by the conjoint action of ER $\alpha$ , FOXA1, and GATA3 to a ligand-inducible growth state. *Mol Syst Biol.* 2011;7(1):526. <https://doi.org/10.1038/msb.2011.59>.
- Guertin MJ, Zhang X, Coonrod SA, Hager GL. Transient estrogen receptor binding and p300 redistribution support a squelching mechanism for estradiol-repressed genes. *Mol Endocrinol.* 2014;28(9):1522–33. <https://doi.org/10.1210/me.2014-1130>.
- Franco HL, Nagari A, Kraus WL. TNF $\alpha$  signaling exposes latent estrogen receptor binding sites to alter the breast cancer cell transcriptome. *Mol Cell.* 2015;58(1):21–34. <https://doi.org/10.1016/j.molcel.2015.02.001>.
- Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. *Nat Protoc.* 2012;7(9):1728–40. <https://doi.org/10.1038/nprot.2012.101>.
- Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics.* 2015;31(14):2382–3. <https://doi.org/10.1093/bioinformatics/btv145>.
- Jalili V, Matteucci M, Masseroli M, Morelli MJ. Using combined evidence from replicates to evaluate ChIP-seq peaks. *Bioinformatics.* 2015;31(17):2761–9. <https://doi.org/10.1093/bioinformatics/btv293>.
- Stark R, Brown G. DiffBind: differential binding analysis of ChIP-Seq peak data. R package version. 2011;1:00(4.3). <https://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf>.
- Tarazona S, Furió-Tari P, Ferrer A, Conesa A. NOISeq: Differential Expression in RNA-seq. 2013. <https://doi.org/10.1093/nar/gkv711>.
- Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 2010;11(3):1–9. <https://doi.org/10.1186/gb-2010-11-3-r25>.
- Tsuyuzaki K, Nikaido I. metaSeq: Meta-analysis of RNA-seq count data. Tokyo University of Science, Tokyo. 2013. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.401.1584&rep=rep1&type=pdf>.

29. Li S, Wan C, Zheng R, Fan J, Dong X, Meyer CA, et al. Cistrome-GO: a web server for functional enrichment analysis of transcription factor ChIP-seq peaks. *Nucleic Acids Res.* 2019;47(W1):W206–11. <https://doi.org/10.1093/nar/gkz332>.
30. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30. <https://doi.org/10.1093/nar/28.1.27>.
31. Keenan AB, Torre D, Lachmann A, Leong AK, Wojciechowicz ML, Utti V, et al. ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Res.* 2019;47(W1):W212–24. <https://doi.org/10.1093/nar/gkz446>.
32. Piryaei Z, Salehi Z, Tahsili MR, Ebrahimie E, Ebrahimi M, Kavousi K. Agonist/antagonist compounds' mechanism of action on estrogen receptor-positive breast cancer: A system-level investigation assisted by meta-analysis. *Inform Med Unlocked.* 2022;31:100985. <https://doi.org/10.1016/j.jimu.2022.100985>.
33. Kumar U, Ardasheva A, Mahmud Z, Coombes RC, Yagüe E. FOXA1 is a determinant of drug resistance in breast cancer cells. *Breast Cancer Res Treat.* 2021;186(2):317–26. <https://doi.org/10.1007/s10549-020-06068-5>.
34. Mohammed H, Russell IA, Stark R, Rueda OM, Hickey TE, Tarulli GA, et al. Progesterone receptor modulates ER $\alpha$  action in breast cancer. *Nature.* 2015;523(7560):313. <https://doi.org/10.1038/nature14583>.
35. Rami F, Baradaran A, Kahnmoori MM, Salehi M. Alteration of GLIS3 gene expression pattern in patients with breast cancer. *Adv Biomed Res.* 2016;5:44. <https://doi.org/10.4103/2277-9175.178803>.
36. Ma S, Tang T, Probst G, Konradi A, Jin C, Li F, et al. Transcriptional repression of estrogen receptor alpha by YAP reveals the Hippo pathway as therapeutic target for ER+ breast cancer. *Nat Commun.* 2022;13(1):1–17. <https://doi.org/10.1038/s41467-022-28691-0>.
37. Luk IY, Reehorst CM, Mariadason JM. ELF3, ELF5, EHF and SPDEF transcription factors in tissue homeostasis and cancer. *Molecules.* 2018;23(9):2191. <https://doi.org/10.3390/molecules23092191>.
38. Deng Y-N, Xia Z, Zhang P, Ejaz S, Liang S. Transcription factor RREB1: from target genes towards biological functions. *Int J Biol Sci.* 2020;16(8):1463. <https://doi.org/10.7150/ijbs.40834>.
39. Chen L-L, Zhang Z-j, Yi Z-b, Li J-j. MicroRNA-211-5p suppresses tumour cell proliferation, invasion, migration and metastasis in triple-negative breast cancer by directly targeting SETBP1. *Br J Cancer.* 2017;117(1):78–88. <https://doi.org/10.1038/bjc.2017.150>.
40. Honkela A, Peltonen J, Topa H, Charapitsa I, Matarese F, Grote K, et al. Genome-wide modeling of transcription kinetics reveals patterns of RNA production delays. *Proc Natl Acad Sci.* 2015;112(42):13115–20. <https://doi.org/10.1073/pnas.1420404112>.
41. Lubecka K, Kaufman-Szymczyk A, Cebula-Obrzut B, Smolewski P, Szemraj J, Fabianowska-Majewska K. Novel clofarabine-based combinations with polyphenols epigenetically reactivate retinoic acid receptor beta, inhibit cell growth, and induce apoptosis of breast cancer cells. *Int J Mol Sci.* 2018;19(12):3970. <https://doi.org/10.3390/ijms19123970>.
42. Singh B, Kinne HE, Milligan RD, Washburn LJ, Olsen M, Lucci A. Important role of FTO in the survival of rare panresistant triple-negative inflammatory breast cancer cells facing a severe metabolic challenge. *PLoS One.* 2016;11(7):e0159072. <https://doi.org/10.1371/journal.pone.0159072>.
43. Hafez AM, Harb OA, Mandour D, Baiomy TA, Eltokhy EA. Clinical, pathological, and prognostic values of chromosomal regional maintenance protein-1 (CRM1) and zinc fingers and homeoboxes-3 (ZHX3) expression in triple-negative breast cancer (TNBC): an immunohistochemical study. *Egypt J Pathol.* 2020;40(1):94. [https://doi.org/10.4103/EJ.PEG.JP\\_36\\_20](https://doi.org/10.4103/EJ.PEG.JP_36_20).
44. Li W, Sang M, Hao X, Xia L, Wang Y, Shan B. Gene expression and DNA methylation analyses suggest that immune process-related ADCY6 is a prognostic factor of luminal-like breast cancer. *J Cell Biochem.* 2020;121(7):3537–46. <https://doi.org/10.1002/jcb.29633>.
45. Pilato B, Pinto R, De Summa S, Lambo R, Paradiso A, Tommasi S. HOX gene methylation status analysis in patients with hereditary breast cancer. *J Hum Genet.* 2013;58(1):51–3. <https://doi.org/10.1038/jhg.2012.118>.
46. Zhao B, Zhou R, Ji C, Liu D, Wu T, Xu H, et al. The function of circRNA-0047604 in regulating the tumor suppressor gene DACH1 in breast cancer. *BioMed Res Int.* 2022;2022:6589651. <https://doi.org/10.1155/2022/6589651>.
47. Asch-Kendrick RJ, Samols MA, Lilo MT, Subhawong AP, Sharma R, Illei PB, et al. NKX3. 1 is expressed in ER-positive and AR-positive primary breast carcinomas. *J Clin Pathol.* 2014;67(9):768–71. <https://doi.org/10.1136/jclinpath-2014-202272>.
48. Denard B, Jiang S, Peng Y, Ye J. CREB3L1 as a potential biomarker predicting response of triple negative breast cancer to doxorubicin-based chemotherapy. *BMC Cancer.* 2018;18(1):1–7. <https://doi.org/10.1186/s12885-018-4724-8>.
49. Corrêa S, Panis C, Binato R, Herrera AC, Pizzatti L, Abdelhay E. Identifying potential markers in breast cancer subtypes using plasma label-free proteomics. *J Proteomics.* 2017;151:33–42. <https://doi.org/10.1016/j.jprot.2016.07.030>.
50. Xu J, Zheng J, Wang J, Shao J. miR-876-5p suppresses breast cancer progression through targeting TFAP2A. *Exp Ther Med.* 2019;18(2):1458–64. <https://doi.org/10.3892/etm.2019.7689>.
51. Song Y, Tian T, Fu X, Wang W, Li S, Shi T, et al. GATA6 is overexpressed in breast cancer and promotes breast cancer cell epithelial–mesenchymal transition by upregulating slug expression. *Exp Mol Pathol.* 2015;99(3):617–27. <https://doi.org/10.1016/j.yexmp.2015.10.005>.
52. Khan J, Masood A, Noor A, Qadir MI. Drug development for PCGF2 protein, involved in formation of human breast cancer by polyphenols and their derivatives extracted from *Cowanina maxicana*. *Chemistry.* 2017;3(6):77–81. <https://doi.org/10.11648/j.ijpc.20170306.12>.
53. Love M, Anders S, Huber W. Differential analysis of count data—the DESeq2 package. *Genome Biol.* 2014;15(550):10–1186. <https://doi.org/10.1186/s13059-014-0550-8>.
54. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139–40. <https://doi.org/10.1093/bioinformatics/btp616>.
55. Marot G, Jaffrézic F, Rau A. metaRNASeq: Differential meta-analysis of RNA-seq data. *BMC Bioinformatics.* 2020;1(26408):3 (<http://cdimage.debian.org/mirror/CRAN/web/packages/metaRNASeq/vignettes/metaRNASeq.pdf>).
56. Moulos P, Hatzis P. Systematic integration of RNA-Seq statistical algorithms for accurate detection of differential gene expression patterns. *Nucleic Acids Res.* 2015;43(4):e25–e. <https://doi.org/10.1093/nar/gku1273>.
57. Zhao F-Q, Misra Y, Li D-B, Wadsworth MP, Krag D, Weaver D, et al. Differential expression of Oct3/4 in human breast cancer and normal tissues. *Int J Oncol.* 2018;52(6):2069–78. <https://doi.org/10.3892/ijo.2018.4341>.
58. Razin S, Borunova V, Maksimenko O, Kantidze O. Cys2His2 zinc finger protein family: classification, functions, and major members. *Biochem Mosc.* 2012;77(3):217–26. <https://doi.org/10.1134/S0006297912030017>.
59. Safran M, Rosen N, Twik M, BarShir R, Stein TI, Dahary D, et al. Practical guide to life science databases. Springer Singapore; 2022. [https://doi.org/10.1007/978-981-16-5812-9\\_2](https://doi.org/10.1007/978-981-16-5812-9_2).
60. Martin S, Lebot M, Sukkarn B, Ball G, Green A, Rakha E, et al. Low expression of G protein-coupled oestrogen receptor 1 (GPER) is associated with adverse survival of breast cancer patients. *Oncotarget.* 2018;9:25946–56. <https://doi.org/10.18632/oncotarget.25408>.
61. Zaman MS, Barman SK, Corley SM, Wilkins MR, Malladi CS, Wu MJ. Transcriptomic insights into the zinc homeostasis of MCF-7 breast cancer cells via next-generation RNA sequencing. *Metallomics.* 2021;13(6):mfab026. <https://doi.org/10.1093/mtomcs/mfab026>.
62. Lan X, Ren R, Feng R, Ly LC, Lan Y, Zhang Z, et al. ZNF410 uniquely activates the NuRD component CHD4 to silence fetal hemoglobin expression. *Mol Cell.* 2021;81(2):239–54. e8. <https://doi.org/10.1016/j.molcel.2020.11.006>.
63. Ma C, Wang F, Han B, Zhong X, Si F, Ye J, et al. SALL1 functions as a tumor suppressor in breast cancer by regulating cancer cell senescence and metastasis through the NuRD complex. *Mol Cancer.* 2018;17(1):1–21. <https://doi.org/10.1186/s12943-018-0824-y>.
64. Basta J, Rauchman M. The nucleosome remodeling and deacetylase complex in development and disease. *Translating epigenetics to the clinic.* 2017:37–72. <https://doi.org/10.1016/B978-0-12-800802-7.00003-4>.
65. Gao X, Liu Z, Zhong M, Wu K, Zhang Y, Wang H, et al. Knockdown of DNA/RNA-binding protein KIN17 promotes apoptosis of triple-negative breast cancer cells. *Oncol Lett.* 2019;17(1):288–93. <https://doi.org/10.3892/ol.2018.9597>.

66. Lin J-H, Lee W-J, Wu H-C, Wu C-H, Chen L-C, Huang C-C, et al. Small G protein signalling modulator 2 (SGSM2) is involved in oestrogen receptor-positive breast cancer metastasis through enhancement of migratory cell adhesion via interaction with E-cadherin. *Cell Adh Migr*. 2019;13(1):121–38. <https://doi.org/10.1080/19336918.2019.1568139>.
67. Guo B-H, Zhang X, Zhang H-Z, Lin H-L, Feng Y, Shao J-Y, et al. Low expression of Mel-18 predicts poor prognosis in patients with breast cancer. *Ann Oncol*. 2010;21(12):2361–9. <https://doi.org/10.1093/annonc/mdq241>.
68. Tan S, Spear R, Zhao J, Sun X, Wang P. Comprehensive characterization of a novel E3-related genes signature with implications in prognosis and immunotherapy of low-grade gliomas. *Front Genet*. 2022;13:905047. <https://doi.org/10.3389/fgene.2022.905047>.
69. Yu D-D, Guo S-W, Jing Y-Y, Dong Y-L, Wei L-X. A review on hepatocyte nuclear factor-1beta and tumor. *Cell Biosci*. 2015;5(1):1–8. <https://doi.org/10.1186/s13578-015-0049-3>.
70. Cai Q, Zhang B, Sung H, Low S-K, Kweon S-S, Lu W, et al. Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. *Nat Genet*. 2014;46(8):886–90. <https://doi.org/10.1038/ng.3041>.
71. Kang I, Chu CT, Kaufman BA. The mitochondrial transcription factor TFAM in neurodegeneration: emerging evidence and mechanisms. *FEBS Lett*. 2018;592(5):793–811. <https://doi.org/10.1002/1873-3468.12989>.
72. Gao W, Wu M, Wang N, Zhang Y, Hua J, Tang G, et al. Increased expression of mitochondrial transcription factor A and nuclear respiratory factor-1 predicts a poor clinical outcome of breast cancer. *Oncol Lett*. 2018;15(2):1449–58. <https://doi.org/10.3892/ol.2017.7487>.
73. Ramachandran A, Basu U, Sultana S, Nandakumar D, Patel SS. Human mitochondrial transcription factors TFAM and TFB2M work synergistically in promoter melting during transcription initiation. *Nucleic Acids Res*. 2017;45(2):861–74. <https://doi.org/10.1093/nar/gkw1157>.
74. Velasco-Ruiz A, Nuñez-Torres R, Pita G, Wildiers H, Lambrechts D, Hatse S, et al. POLRMT as a novel susceptibility gene for cardiotoxicity in epirubicin treatment of breast cancer patients. *Pharmaceutics*. 2021;13(11):1942. <https://doi.org/10.3390/pharmaceutics13111942>.
75. Rakha EA, Alsaleem M, ElSharawy KA, Toss MS, Raafat S, Mihai R, et al. Visual histological assessment of morphological features reflects the underlying molecular profile in invasive breast cancer: a morphomolecular study. *Histopathology*. 2020;77(4):631–45. <https://doi.org/10.1111/his.14199>.
76. Estruch IM. Novel insights in the molecular mechanisms of action of retinoids and their potential repercussions on breast cancer cell proliferation: Wageningen University and Research. 2018. <https://doi.org/10.18174/448553>.
77. Williams KE, Jawale RM, Schneider SS, Otis CN, Pentecost BT, Arcaro KF. DNA methylation in breast cancers: Differences based on estrogen receptor status and recurrence. *J Cell Biochem*. 2019;120(1):738–55. <https://doi.org/10.1002/jcb.27431>.
78. Wang Q, Liu X. Screening of feature genes in distinguishing different types of breast cancer using support vector machine. *Onco Targets Ther*. 2015;8:2311. <https://doi.org/10.2147/OTT.S85271>.
79. Fang L, Wang Y, Gao Y, Chen X. Overexpression of CXXC5 is a strong poor prognostic factor in ER+ breast cancer. *Oncol Lett*. 2018;16(1):395–401. <https://doi.org/10.3892/ol.2018.8647>.
80. Yan LR, Wang A, Lv Z, Yuan Y, Xu Q. Mitochondria-related core genes and TF-miRNA-hub mrDEGs network in breast cancer. *Biosci Rep*. 2021;41(1):BSR20203481. <https://doi.org/10.1042/BSR20203481>.
81. Yin Y, Guan Z, Ou Z, Tang M, Li Y, Zhang C, et al. For a systematic comprehension of GAB2: an analysis based on pan-cancer. 2022. <https://doi.org/10.21203/rs.3.rs-1224365/v1>.
82. Pils D, Hager G, Tong D, Aust S, Heinze G, Kohl M, et al. Validating the impact of a molecular subtype in ovarian cancer on outcomes: a study of the OVCAD Consortium. *Cancer Sci*. 2012;103(7):1334–41. <https://doi.org/10.1111/j.13497006.2012.02306.x>.
83. Ho XD, Phung P, Le VQ, H Nguyen V, Reimann E, Prans E, et al. Whole transcriptome analysis identifies differentially regulated networks between osteosarcoma and normal bone samples. *Exp Biol Med*. 2017;242(18):1802–11.
84. Mashima T, Tameda Y, Jang M-K, Mizutani A, Muramatsu Y, Yoshida H, et al. mTOR signaling mediates resistance to tankyrase inhibitors in Wnt-driven colorectal cancer. *Oncotarget*. 2017;8(29):47902. <https://doi.org/10.18632/oncotarget.18146>.
85. Fu J, Zhang X, Yan L, Shao Y, Liu X, Chu Y, et al. Identification of the hub gene BUB1B in hepatocellular carcinoma via bioinformatic analysis and in vitro experiments. *PeerJ*. 2021;9:e10943. <https://doi.org/10.7717/peerj.10943>.
86. Alsagaby SA, Vijayakumar R, Premanathan M, Mickymaray S, Alturaiki W, Al-Baradie RS, et al. Transcriptomics-based characterization of the toxicity of ZnO nanoparticles against chronic myeloid leukemia cells. *Int J Nanomed*. 2020;15:7901. <https://doi.org/10.2147/IJN.S261636>.
87. Zhou Z, Bai J, Zhong S, Zhang R, Kang K, Zhang X, et al. Integrative Functional Genomic Analysis of Molecular Signatures and Mechanistic Pathways in the Cell Cycle Underlying Alzheimer's Disease. *Oxid Med Cell Longev*. 2021;2021:5552623. <https://doi.org/10.1155/2021/5552623>.
88. Erdogan C, Suer I, Kaya M, Kurt Z, Ozturk S, Aydin N. Bioinformatics Analysis of the Potentially Functional circRNA-miRNA-mRNA Network in Breast Cancer. *bioRxiv*. 2022. <https://doi.org/10.1101/2022.01.10.475557>.
89. Chen B, Tang H, Chen X, Zhang G, Wang Y, Xie X, et al. Transcriptomic analyses identify key differentially expressed genes and clinical outcomes between triple-negative and non-triple-negative breast cancer. *Cancer Manag Res*. 2019;11:179. <https://doi.org/10.2147/CMAR.S187151>.
90. Murugesan M, Premkumar K. Integrative miRNA-mRNA functional analysis identifies miR-182 as a potential prognostic biomarker in breast cancer. *Molecular Omics*. 2021;17(4):533–43. <https://doi.org/10.1039/d0mo00160k>.
91. Wu Y, Luo S, Yin X, He D, Liu J, Yue Z, et al. Co-expression of key gene modules and pathways of human breast cancer cell lines. *Biosci Rep*. 2019;39(7):BSR20181925. <https://doi.org/10.1042/BSR20181925>.
92. Guo Q, Pei XH, Chu AJ, Guo YB, Fan YY, Wang CH, et al. The mechanism of action of Fangji Huangqi Decoction on epithelial-mesenchymal transition in breast cancer using high-throughput next-generation sequencing and network pharmacology. *J Ethnopharmacol*. 2022;284:114793. <https://doi.org/10.1016/j.jep.2021.114793>.
93. Zhang M, Wu K, Wang M, Bai F, Chen H. CASP9 As a Prognostic Biomarker and Promising Drug Target Plays a Pivotal Role in Inflammatory Breast Cancer. 2021. <https://doi.org/10.21203/rs.3.rs-355512/v1>.
94. Yang Q, Li K, Li X, Liu J. Identification of key genes and pathways in myeloma side population cells by bioinformatics analysis. *Int J Med Sci*. 2020;17(14):2063. <https://doi.org/10.7150/ijms.48244>.
95. Cheng R, Qi L, Kong X, Wang Z, Fang Y, Wang J. Identification of the Significant Genes Regulated by Estrogen Receptor in Estrogen Receptor-Positive Breast Cancer and Their Expression Pattern Changes When Tamoxifen or Fulvestrant Resistance Occurs. *Front Genet*. 2020;11. <https://doi.org/10.3389/fgene.2020.538734>.
96. Li D, Liu D, Yue D, Gao P, Du C, Liu X, et al. Network pharmacology and RNA sequencing studies on triterpenoid saponins from *Bupleurum chinense* for the treatment of breast cancer. *RSC Adv*. 2019;9(70):41088–98. <https://doi.org/10.1039/C9RA08970E>.
97. Kiliç N, Islakoğlu YÖ, Büyükl İ, Gür-Dedeoğlu B, Cansaran-Duman D. Determination of usnic acid responsive miRNAs in breast cancer cell lines. *Anticancer Agents Med Chem*. 2019;19(12):1463–72. <https://doi.org/10.2174/1871520618666181112120142>.
98. Pyrczak-Felczykowska A, Reekie TA, Jąkowski M, Hać A, Malinowska M, Pawlik A, et al. The isoxazole derivative of usnic acid induces an ER stress response in breast cancer cells that leads to paraptosis-like cell death. *Int J Mol Sci*. 2022;23(3):1802.
99. Chen G-Q, Zhao Z-W, Zhou H-Y, Liu Y-J, Yang H-J. Systematic analysis of microRNA involved in resistance of the MCF-7 human breast cancer cell to doxorubicin. *Med Oncol*. 2010;27(2):406–15. <https://doi.org/10.1007/s12032-009-9225-9>.
100. Wang J, Zhang Q, Zhou S, Xu H, Wang D, Feng J, et al. Circular RNA expression in exosomes derived from breast cancer cells and patients. *Epigenomics*. 2019;11(4):411–21. <https://doi.org/10.2217/epi-2018-0111>.
101. Liu S, Hu X, Fan X, Jin R, Yang W, Geng Y, et al. A Bioinformatics research on novel mechanism of compound kushen injection for treating breast cancer by network pharmacology and molecular docking verification. *Evid Based Complement Alternat Med*. 2020;2020:2758640. <https://doi.org/10.1155/2020/2758640>.

102. Wang J, Wang X, Zhang H, Lu M, Niu X, Yin L, et al. Analyzing the multi-target pharmacological mechanism of folium artemisiae argyi acting on breast cancer: a network pharmacology approach. *Ann Transl Med.* 2020;10(24):1368. <https://doi.org/10.21203/rs.3.rs-127242/v1>.
103. Ray M, Sarkar S. In silico identification of novel non-synonymous variants in metabolic pathway associated target genes of papillary thyroid carcinoma: A way towards future treatment of papillary thyroid carcinoma. *Meta Gene.* 2020;24:100700. <https://doi.org/10.1016/j.mgene.2020.100700>.
104. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature.* 2012;481(7381):389–93. <https://doi.org/10.1038/nature10730>.
105. Ye T, Feng J, Wan X, Xie D, Liu J. Double agent: SPDEF gene with both oncogenic and tumor-suppressor functions in breast cancer. *Cancer Manag Res.* 2020;12:3891. <https://doi.org/10.2147/CMAR.S243748>.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

