Genetic and epigenetic modifications of *HPDL* and *SOX17* associated with breast cancer prognosis: a study based on The Cancer Genome Atlas

Chundi Gao
Shandong University of Traditional Chinese Medicine
Huayao Li
Shandong University of Traditional Chinese Medicine
Cun Liu
Shandong University of Traditional Chinese Medicine
Jibiao Wu
Shandong University of Traditional Chinese Medicine
Chao Zhou
Weifang Traditional Chinese Hospital
Lijuan Liu
Weifang Traditional Chinese Hospital
Jing Zhuang
Weifang Traditional Chinese Hospital
Changgang sun (scgdoctor@126.com)
Department of Oncology,Affilited Hospital of Shandong University of Traditional Chinese Medicine
https://orcid.org/0000-0002-6648-3602

Primary research

Keywords: breast cancer, methylation, copy number variation, multi-layer correlation analysis, survival analysis

Posted Date: March 25th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-18876/v1

Abstract

Background The high heterogeneity of breast cancer (BRCA) makes it more challenging to interpret the genetic variation mechanisms involved in BRCA pathogenesis and prognosis. Areas with high DNA methylation (such as CpG islands) were accompanied by copy number variation (CNV), and these genomic variations affected the level of DNA methylation.

Methods: In this study, we characterized inter-tumor heterogeneity and analyzed the effects of CNV on DNA methylation and gene expression. In addition, we performed a Genetic Set Enrichment Analysis (GSEA) to identify key pathways for changes between patients with low and high expression of genes.

Results: Our analysis found that the CNV of *HPDL* and *SOX17* is not only related to the patient's prognosis, but also related to gene methylation and expression levels affecting the patient's survival time.

Conclusion: This study provided an effective bioinformatics basis for further exploration of molecular mechanisms related to BRCA and assessment of patient prognosis, but the development of biomarkers for diagnosis and treatment still requires further clinical data validation.

Background

In the post-genome era, rapidly evolving high-throughput sequencing technologies have enabled the acquisition of vast amounts of multi-omics data more efficiently [1]. The variation of expression of some genes causes the genetic regulation trajectory inside the cell to deviate, which alters the gene expression programming inside the cell. Therefore, most disease-causing genomic variants are likely to play a role by altering gene regulation, such as transcription factor binding and DNA methylation, rather than directly affecting protein function [2, 3]. The high heterogeneity of breast cancer (BRCA) makes it more challenging to interpret the genetic variation mechanisms involved in BRCA pathogenesis and prognosis [4].

In human cancer, genomic instability leads to extensive cell copy number variation (CNV) [5]. Genome-wide association studies (GWASs) have been conducted for common malignancies and have identified more than 450 genetic variants associated with increased disease risk [6]. Budczies J et al. found that PD-L1 CNVs were significantly associated with changes in PD-L1 mRNA expression in 22 cancer types; Tumors with increased PD-L1 expression exhibited significantly higher mutation loads, and could be, therefore, used in immunotherapeutic regimens and as predictive biomarkers [7]. In BRCA, CNV is associated with the expression of ~ 40% of genes, and thus, participate in the onset, treatment, and recurrence of BRCA, and affect the prognosis of BRCA patients [8]. CNV in BRCA1, BLM, OR4C11, OR4P4, CDH1, HEPACAM, and LOXHD1 increase the incidence of BRCA [9, 10]. MCL1, MYC and JAK2, and PTEN deletions or mutations play a role in de novo or acquired chemotherapeutic resistance in triple-negative BRCA [11]. The CNV of FGFR1 and ZNF703 increase the risk of death in patients with BRCA [12, 13]. Higher intratumoral heterogeneity of EGFR/CEP7 and CCND1/CEP11 CNVs predicts metastasis and is associated with significantly worse metastasis-free survival in triple-negative BRCA patients [14].

Disorders in the epigenetic state are closely related to human diseases, particularly cancer. DNA methylation is a well-characterized epigenetic modification that is closely related to many cellular processes. In the current study, DNA methylation and its sites associated with tumor recurrence and overall survival (OS) of BRCA and its subtypes have been identified based on methods employed for genome-wide DNA methylation analysis [15–17]. The methylation of oncogenes, ESR1 and ERBB2, and tumor suppressor genes, FBLN2, CEBPA, and FAT4, contribute to the early diagnosis of BRCA [18]. Methylation of HER2, Ki67, and GSTP1 are associated with BRCA TNM staging and tumor

size and can be combined for early diagnosis and prognosis [19]. FECR1 circular RNA expression is coordinated by regulation of DNA methylation and demethylation. Upstream regulators control BRCA tumor growth [20]. However, MLH1/MSH2, methylation product of the DNMT gene may be important for the chemotherapeutic tolerance of BRCA [21, 22].

CNV represents a major source of genomic variation and is an important genetic factor leading to various cancers. DNA methylation, a major means of epigenetic modification, is considered an inhibitory epigenetic marker. Several studies have found that areas with high DNA methylation (such as CpG islands) are accompanied by copy number variation, and these genomic variations affect the level of DNA methylation [23]. For example, in lung adenocarcinoma, DNA methylation heterogeneity demonstrates branch clonal evolution of lung adenocarcinoma regions driven by genomic instability, and subclone copy number variation [24]. Here, we investigated the association between genomic variation (such as CNV) in regulatory regions of BRCA and corresponding changes in DNA methylation. In addition, we performed a Genetic Set Enrichment Analysis (GSEA) to identify key pathways for changes between patients with low and high expression of genes. Thus, an in-depth study of the genome pathogenesis of BRCA was conducted to identify prognostic biomarkers and their clinical efficiency.

Our analysis found that HPDL and SOX17, tumor suppressor genes, can cause malignant transformation of cells and cause tumorigenesis when mutations, deletions, or inactivations occur. In addition, the results of the study indicate that the CNV of HPDL and SOX17 is not only related to the patient's prognosis, but also related to gene methylation and expression levels affecting the patient's survival time.

Materials And Methods

Data processing and analysis

TCGA Data Portal was terminated, and all TCGA data were transferred to the Genomic Data Commons (GDC). Therefore, data from The Cancer Genome Atlas (TCGA) of BRCA-related methylation, CNV, gene expression, and clinical data were downloaded from GDC (https://gdc.cancer.gov/). The chi-square test, and Limma and Edger software packages were used to collate and analyze the downloaded data and screened according to P and logFC values. To obtain differences in CNV, abnormally methylated and dysregulated genes between BRCA tissue samples and normal tissue samples were analyzed. The data from the TCGA database is public. Therefore, no approval from the local ethics committee was required.

Multi-layer correlation analysis method predicts the pattern of gene CNV in BRCA

DNA methylation has been shown to regulate gene expression in a variety of ways, such as changing chromosome structure, DNA stability, etc. In addition, CNV is widely distributed in the human genome and has important biological implications. To further explore the link between CNV and methylation on gene expression, the possible patterns of CNV in BRCA need to be elucidated. This study focuses on the analysis of correlation between abnormal methylation and gene expression, CNV and aberrant methylation, and CNV and gene expression. Screening was done by the Pearson correlation coefficient and p-value. Key genes with simultaneous methylation abnormalities, CNV, and abnormal expression were obtained, and further prognostic analysis was performed on these genes.

Mapping of Kaplan-Meier survival curve of genes and screening of prognostic key genes

In order to further identify key genes related to the prognosis of BRCA patients from the genes obtained above, survival analysis was performed on the relevant data based on the survival software package, and survival curves

were plotted to show the effect of abnormal methylation and methylation combined with abnormal gene expression on patient survival. In addition, in order to further explore the methylation sites of prognostic aberrant methylation genes, the factors affecting the prognosis of patients and gene expression are mapped to specific methylation sites.
Effect Of Key Genes Cnv On Patient Prognosis

Through data analysis, it was found that the abnormal methylation of key genes is closely related to the prognosis of BRCA patients, while the key genes harbored methylation abnormalities, CNV, and abnormal expression, and there was a significant correlation between them. The effect of mutations on the prognosis of patients can be seen by studying CNV and survival time of BRCA patients, further indicating the biological significance of gene CNV in the progression of BRCA. In addition, we performed a gene set enrichment analysis (GSEA) analysis between high-expression and low-expression groups of key genes to determine key pathways that are altered in patients with abnormal gene expression [25].

Results

Data processing and analysis

In this study, BRCA-related methylation data downloaded from the TCGA database included 883 samples, comprising 96 normal tissue samples and 787 BRCA tissue samples. The difference analysis results obtained a total of 122 protein-coding genes with $P < 0.05$ and $|logFC| > 1$ as the cut-off condition (Fig. 1A). The CNV data included 2201 samples, 1103 normal tissue samples, and 1098 BRCA tissue samples. A total of 19178 genes with CNV were found based on the chi-square test results ($P < 0.05$) (Supplementary material 1). The difference analysis of gene expression data between 112 normal tissue samples and 1096 cancer tissue samples showed that 2138 dysregulated genes, including 1375 upregulated genes and 763 downregulated genes (Fig. 1B), were obtained with $P < 0.01$ and $|logFC| > 2$ as the cut-off condition.
Multi-layer Correlation Analysis Method To Screen Key Genes

In order to reduce the number of calculations of correlation analysis between the two, we performed correlation analysis on the condition of genes with abnormal methylation. First, we found that 105 of the 122 genes with aberrant methylation exhibited simultaneous expression disorders. Correlation analysis showed that the aberrant methylation of 25 genes was closely related to the expression with the Pearson correlation coefficient $Cor > 0.4$ as the screening criterion (Table 1). Interestingly, these 25 genes harbored CNV simultaneously (Fig. 1C). To explore the pattern of effects of CNV in disease progression, we performed a correlation analysis of CNV with methylationand abnormal gene expression for 25 genes. Screening with $P < 0.01$ as the cut off criterion, the CNV of 12 genes was associated with the level of methylation, and the CNV of 16 genes was related to the abnormal expression level. Among them, there are 6 common genes. (Fig. 2, Supplementary material 2A,2B). We used these six genes as key genes for prognostic survival analysis.

Table 1
25 genes with significant correlation between methylation level and expression

Gene	Cor	Pvalue
HOXB13	0.455	5.47E-45
SCGB3A1	-0.404	5.75E-35
POU4F1	-0.411	2.76E-36
SOX17	-0.412	2.06E-36
SLC35G2	-0.417	2.56E-37
RASSF10	-0.421	4.45E-38
AADAT	-0.422	2.84E-38
HFM1	-0.424	1.23E-38
HPDL	-0.426	3.82E-39
SNCA	-0.427	2.85E-39
TBX18	-0.429	1.18E-39
VIM	-0.475	1.99E-49
DSC3	-0.503	4.94E-56
LRRC34	-0.505	1.23E-56
ZSCAN23	-0.53	2.24E-63
ZNF454	-0.538	2.08E-65
ZNF728	-0.545	1.39E-67
MT1E	-0.566	9.90E-74
IRX1	-0.567	4.62E-74
ZNF492	-0.584	1.55E-79
PSAT1	-0.587	1.64E-80
EID3	-0.599	1.23E-84
ID4	-0.605	1.26E-86
HIST3H2A	-0.658	2.66E-107
ZNF471	-0.662	3.79E-109

Joint survival analysis and site-related prognostic assessment to identify biomarkers

Through joint survival analysis, it was found that the combination of methylation and abnormal expression of HPDL and SOX17 was significantly associated with the prognosis of BRCA patients. Furthermore, the results showed that

high-methylation low-expression of HPDL and SOX17 showed poor prognosis (Fig. 3A). In addition, based on the survival of the R package, we analyzed the effects of the relevant methylation sites of these two genes on patient survival. $P < 0.05$ was used as a screening criterion for predicting prognosis, and specific methylation sites associated with the prognosis of these genes were found. Among them, the two methylation sites of HPDL and the eight methylation sites of SOX17 can affect the survival time of patients (Fig. 3B).

Kaplan-Meier survival curve analysis of the effect of gene CNV on patient prognosis

The genes HPDL and SOX17 showed not only methylation abnormalities and abnormal expression, but also CNV. Further analysis showed that CNV in HPDL and SOX17 were associated with overall patient survival, in which the addition of two copies of SOX17 is associated with a lower survival rate, while a decrease in the copy number of HPDL also suggests a poor prognosis (Fig. 3C). In addition, as the CNV of HPDL and SOX17 are related to methylation and abnormal expression levels, our research indicated that the CNV of HPDL and SOX17 can directly affect the prognosis of patients, and can also indirectly affect the survival time of patients by affecting the methylation and expression levels of the corresponding genes.

GSEA analysis of patients with low and high expression of HPDL and SOX17

To identify the molecular pathways of the biological functions and effects of HPDL and SOX17 in BRCA progression; we used GSEA to identify key pathways involved in the changes between patients with low and high expression of genes. With p value < 0.05 as the screening standard, the results indicated that the pathways that HPDL can affect mainly, included MAPK signaling pathway,p53 signaling pathway, etc. In addition, SOX17 mainly affects JAK STAT signaling pathway, WNT signaling pathway, and so on (Table 2, Fig. 4).

Table 2

The key pathways for the differential between low and high expression of patients based on GSEA analysis.

	NAME	NES	p-val
HPDL	KEGG_FOCAL_ADHESION	-1.91198	0
	KEGG_DILATED_CARDIOMYOPATHY	-1.61559	0.002
	KEGG_HYPERTROPHIC_CARDIOMYOPATHY_HCM	-1.64029	0.002075
	KEGG_ADHERENS_JUNCTION	-1.74735	0.002283
	KEGG_ALDOSTERONE_REGULATED_SODIUM_REABSORPTION	-1.70915	0.002387
	KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY_ARVC	-1.75006	0.006993
	KEGG_ABC_TRANSPORTERS	-1.5166	0.014141
	KEGG_MAPK_SIGNALING_PATHWAY	-1.66251	0.016713
	KEGG_TGF_BETA_SIGNALING_PATHWAY	-1.52333	0.04008
	KEGG_OOCYTE_MEIOSIS	1.754743	0
	KEGG_PROGESTERONE_MEDIATED_OOCYTE_MATURATION	1.685132	0
	KEGG_CELL_CYCLE	1.639552	0
	KEGG_P53_SIGNALING_PATHWAY	1.51856	0.035503
	KEGG_OLFACTORY_TRANSDUCTION	1.609742	0.046472
SOX17	KEGG_OOCYTE_MEIOSIS	-1.58789	0.04814
	KEGG_JAK_STAT_SIGNALING_PATHWAY	1.797007	0
	KEGG_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION	1.865758	0.001481
	KEGG_ABC_TRANSPORTERS	1.455277	0.014463
	KEGG_RETINOL_METABOLISM	1.741166	0.017483
	KEGG_HEMATOPOIETIC_CELL_LINEAGE	1.564893	0.021195
	KEGG_VASCULAR_SMOOTH_MUSCLE_CONTRACTION	1.562236	0.023636
	KEGG_ADIPOCYTOKINE_SIGNALING_PATHWAY	1.588887	0.029297
	KEGG_INSULIN_SIGNALING_PATHWAY	1.672106	0.038113
	KEGG_METABOLISM_OF_XENOBIOTICS_BY_CYTOCHROME_P450	1.663495	0.039587
	KEGG_WNT_SIGNALING_PATHWAY	1.528532	0.04
	KEGG_DRUG_METABOLISM_CYTOCHROME_P450	1.675274	0.043328
	KEGG_AXON_GUIDANCE	1.497392	0.047445

Discussion

Heterogeneity is an important predictor of tumor treatment failure and drug resistance, and genomic mutations (such as copy number variation) are important causal factors of heterogeneity among tumors. Previous studies have shown that CNV can affect the expression level of proteins through epigenetic regulation, and the key mechanism is to affect epigenetic modifications (such as DNA methylation). The overall hypomethylation of oncogenes and hypermethylation of tumor suppressor genes are characteristic of most cancer types. Molecular understanding of BRCA heterogeneity is the key to effective treatment and personalized medicine. In this study, BRCA-related methylation data downloaded from the TCGA database included 883 samples, comprising 96 normal tissue samples and 787 BRCA tissue samples, the difference analysis results obtained a total of 122 protein-coding genes. The CNV data included 2201 samples, 1103 normal tissue samples, and 1098 BRCA tissue samples, a total of 19178 genes with CNV were found. The difference analysis of gene expression data between 112 normal tissue samples and 1096 cancer tissue samples showed that 2138 dysregulated genes, including 1375 upregulated genes and 763 downregulated genes. We used TCGA high-throughput molecular profiling data to characterize inter-tumor heterogeneity and analyzed the effects of CNV on DNA methylation and gene expression.

In our analysis, CNV of HPDL and SOX17 affected methylation and gene expression levels in BRCA, and CNV and methylation of HPDL and SOX17 can lead to poor prognosis in patients with BRCA. The CNV of SOX17 and HPDL can influence the expression of genes through genetic regulation. The CNV of SOX17 showed copy number amplification on chromosome 8, and further analysis showed that the prognosis of BRCA patients was poor when SOX17 copy number was increased. The CNV of HPDL showed a decrease in copy number on chromosome 1, and further analysis indicated that the survival rate of BRCA patients was lower when HPDL copy number was reduced. The CNV of SOX17 and HPDL can affect the expression of genes through epigenetic modification, and DNA methylation is an important pathway for epigenetic modification. The methylation sites of SOX17 that we characterized with BRCA OS included cg00123055, cg02222728, cg03329976, cg08044907, cg15377283, cg24150172, cg24891539, and cg24928317. The methylation sites of HPDL included cg12178578 and cg15071854. Survival analysis showed that the OS of BRCA patients hypermethylated in SOX17 and HPDL was poorer. Therefore, CNV and methylation of SOX17 and HPDL are early events of BRCA and can predictrecurrence, metastasis, and prognosis of BRCA patients.

SOX17, a transcriptional regulator, binds to target promoter DNA and inhibits Wnt signaling. SOX17 gene promoter methylation can be used as a tumor suppressor and dysregulated oncogene (via aberrant DNA methylation) in many tumors, such as lung adenocarcinoma [26], cholangiocarcinoma [27], esophageal squamous cell carcinoma [28], colorectal cancer [29], non-small cell lung cancer [30], endometrial cancer [31], and so on. In BRCA, Fu Deyuan et al. used methylation-specific polymerase chain reaction to assess the relationship between the methylation of the SOX17gene promoter and the onset and prognosis of BRCA. Abnormal SOX17 methylation in cancer tissues and plasma DNA was found to be significantly associated with tumor lymph node metastasis and lymph node metastasis, associated with poor disease-free survival ($P < 0.005$) and overall survival ($P < 0.005$). In addition, SOX17 methylation in plasma DNA is an independent prognostic factor for DFS in BRCA [32]. Chimonidou Maria et al. found that the SOX17 promoter is highly methylated in primary breast tumors, in CTCs isolated from patients with BRCA, and in corresponding cfDNA samples, which provides new predictive ideas for recurrence and prognosis in patients with operable BRCA and metastatic patients [33, 34]. HPDL may have dioxygenase activity. Previous studies have found that HPDL exhibits differential expression in CNS lymphoma compared with non-primary central nervous system (CNS) lymphoma [35]. However, understanding the role of HPDL in BRCA needs further research and interpretation, which provides an idea for the in-depth study of the molecular mechanism of BRCA.

Intracellular signaling pathways regulate various cellular activities. We performed GSEA identification on SOX17 and HPDL to further explore the small molecule regulation mechanism of BRCA and found that signaling pathways with significant changes in enrichment exist between patients with low expression and high expression. It is well known

that the JAK-STAT signaling pathway, a signal transduction pathway stimulated by cytokines, is involved in biological processes, such as cell proliferation, differentiation, apoptosis, and immune regulation, and is associated with pathogenesis of many tumors, such as liver cancer, ovarian cancer, and BRCA [36–38].The major cellular processes during BRCA development rely on JAK/STAT signaling to coordinate growth factor function. Previous studies have found that activation of the JAK/STAT pathway is common in triple-negative BRCA, which can affect the expression of genes controlling immune signals. Dysregulated JAK/STAT signaling has been implicated in BRCA metastasis, associated with high risk of recurrence [39–41]. The Wnt signaling pathway plays a crucial role in early embryonic development, organ formation, tissue regeneration, and other physiological processes, often involving stem cell control, which may induce cancer if a key protein is mutated [42]. Wnt signaling pathway involves the onset and treatment of colorectal cancer, pancreatic cancer, gastric cancer, and other tumors [43–45]. Yang Feibiao et al. confirmed that SOX17 is a target gene of miR-194-5p. In mouse studies, knockdown of miR-194-5p in BRCA cells may increase SOX17 expression and regulate the signaling pathway of Wnt/β-catenin [46]. Therefore, increased expression of SOX17 is associated with activation of the Wnt signaling pathway and is thus involved in the pathogenesis of BRCA. In addition, the enrichment results of SOX17 include pathways related to cell growth, division, and proliferation of oocyte meiosis, ABC transporters, and neuroactive ligand receptor interaction. HPDL upregulation is related to cell cycle and P53 signaling pathway. HPDL down-regulation is related to MAPK signaling pathway and TGF-β signaling pathway. Both cell cycle and p53 signaling pathways are involved in cell division and proliferation. The p53 gene is called the “guardian of the genome”, but when p53 is deregulated, it participates in the development and proliferation of various tumor cells [47].Both MAPK and TGF-βsignaling pathway are involved in cell growth, differentiation, and apoptosis. In recent studies, abnormal activation of the MAPK signaling pathway signal has been found to favor the abnormal proliferation of malignant cells [48]. TGF-βsignaling acts as suppressor and inducer of tumor progression during the early and late stages of cancer, and can trigger a cascade of reactions that mobilize cancer cells [49, 50].

Recent studies have demonstrated the consequences of genetic variation in regulating overall risk associated with BRCA patients. In the study so far, we explored the effects of CNV and DNA methylation on gene expression levels and OS of BRCA patients and found that CNV can affect DNA methylation levels. CNV and methylation of SOX17 and HPDL are related to expression and regulation. In addition, the CNV of SOX17 and HPDL were also correlated with methylation levels. In addition, we found methylation sites for SOX17 and HPDL associated with BRCA prognosis. DNA methylation is an effective regulator of gene expression, If the CpG island is located in the promoter region of a gene, the methylation of the CpG island will significantly reduce or even completely silence the transcription of the gene, and then affect the protein expression. In this study, due to data and conditional restrictions, we did not distinguish whether it was on the promoter or DNA when screening prognostic related methylation sites, which is what we will explore in the next study. Finally, by enriching the low and high expression pathways of SOX17 and HPDL, pathways related to BRCA progression have been discovered, including the JAK-STAT / Wnt / P53 / MAPK signaling pathway.

Conclussion

In summary, by comprehensively assessing the effects of CNV and DNA methylation on gene expression and patient OS, the CNV and DNA methylation associated with risk of BRCA recurrence and prognosis were identified. These new discoveries are very promising. Prognostic assessment at the genome level may not only be useful for identifying new prognostic biomarkers, but would also open up new horizons for novel pathways involved in BRCA progression, serving the potential goal of developing more effective therapeutic strategies.

Abbreviations

CNV Copy number variation TCGA: The Cancer Genome Atlas; GSEA : Genetic Set Enrichment Analysis OS :Overall Survival.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable

Availability of data and materials

All data generated or analysed during this study are included in this published article.

Competing interests

The authors declare that they have no competing interests.

Funding

This work is supported by the grants from National Natural Science Foundation of China 81673799 and National Natural Science Foundation of China (81703915).

Authors' contributions

CG and CS conceived and designed the study; JZ HL and JW performed data analysis; CL, CZ, LL contributed analysis tools; CG and HL was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Author details

[1]College of First Clinical Medicine, Shandong University of Traditional Chinese Medicine, Jinan, Shandong, PR China. (CG and CL);[2]College of Basic medical, Shandong University of Traditional Chinese Medicine, Jinan, Shandong, PR China. (HL and JW).[3]Departmen of Oncology, Weifang Traditional Chinese Hospital, Weifang, Shandong, PR China. (JZ, CZ and LL);[4]Cancer and Immunology Institute, Shandong University of Traditional Chinese Medicine, Jinan, Shandong Province, P.R. China. (CS).

References

1. Tomczak K, Czerwinska P, Wiznerowicz M: The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemporary oncology (Poznan, Poland)* 2015, 19(1a):A68-77.
2. Kilpinen H, Waszak SM, Gschwind AR, Raghav SK, Witwicki RM, Orioli A, Migliavacca E, Wiederkehr M, Gutierrez-Arcelus M, Panousis NI *et al*: Coordinated effects of sequence variation on DNA binding, chromatin structure, and

transcription. *Science (New York, NY)* 2013, 342(6159):744-747.

3. Schubeler D: Function and information content of DNA methylation. *Nature* 2015, 517(7534):321-326.
4. Sharma R: Breast cancer incidence, mortality and mortality-to-incidence ratio (MIR) are associated with human development, 1990-2016: evidence from Global Burden of Disease Study 2016. *Breast cancer (Tokyo, Japan)* 2019.
5. Paolella BR, Gibson WJ, Urbanski LM, Alberta JA, Zack TI, Bandopadhayay P, Nichols CA, Agarwalla PK, Brown MS, Lamothe R *et al*: Copy-number and gene dependency analysis reveals partial copy loss of wild-type SF3B1 as a novel cancer vulnerability. *eLife* 2017, 6.
6. Sud A, Kinnersley B, Houlston RS: Genome-wide association studies of cancer: current insights and future perspectives. *Nature reviews Cancer* 2017, 17(11):692-704.
7. Budczies J, Bockmayr M, Denkert C, Klauschen F, Groschel S, Darb-Esfahani S, Pfarr N, Leichsenring J, Onozato ML, Lennerz JK *et al*: Pan-cancer analysis of copy number changes in programmed death-ligand 1 (PD-L1, CD274) - associations with gene expression, mutational load, and survival. *Genes, chromosomes & cancer* 2016, 55(8):626-639.
8. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y *et al*: The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012, 486(7403):346-352.
9. Schulten HJ, Bangash M, Karim S, Dallol A, Hussein D, Merdad A, Al-Thoubaity FK, Al-Maghrabi J, Jamal A, Al-Ghamdi F *et al*: Comprehensive molecular biomarker identification in breast cancer brain metastases. *Journal of translational medicine* 2017, 15(1):269.
10. Walker LC, Pearson JF, Wiggins GA, Giles GG, Hopper JL, Southey MC: Increased genomic burden of germline copy number variants is associated with early onset breast cancer: Australian breast cancer family registry. *Breast cancer research : BCR* 2017, 19(1):30.
11. Wein L, Loi S: Mechanisms of resistance of chemotherapy in early-stage triple negative breast cancer (TNBC). *Breast (Edinburgh, Scotland)* 2017, 34 Suppl 1:S27-s30.
12. Chen H, Singh RR, Lu X, Huo L, Yao H, Aldape K, Abraham R, Virani S, Mehrotra M, Mishra BM *et al*: Genome-wide copy number aberrations and HER2 and FGFR1 alterations in primary breast cancer by molecular inversion probe microarray. *Oncotarget* 2017, 8(7):10845-10857.
13. Holland DG, Burleigh A, Git A, Goldgraben MA, Perez-Mancera PA, Chin SF, Hurtado A, Bruna A, Ali HR, Greenwood W *et al*: ZNF703 is a common Luminal B breast cancer oncogene that differentially regulates luminal and basal progenitors in human mammary epithelium. *EMBO molecular medicine* 2011, 3(3):167-180.
14. Yang F, Wang Y, Li Q, Cao L, Sun Z, Jin J, Fang H, Zhu A, Li Y, Zhang W *et al*: Intratumor heterogeneity predicts metastasis of triple-negative breast cancer. *Carcinogenesis* 2017, 38(9):900-909.
15. Fackler MJ, Umbricht CB, Williams D, Argani P, Cruz LA, Merino VF, Teo WW, Zhang Z, Huang P, Visvananthan K *et al*: Genome-wide methylation analysis identifies genes specific to breast cancer hormone receptor status and risk of recurrence. *Cancer research* 2011, 71(19):6195-6207.
16. Fleischer T, Frigessi A, Johnson KC, Edvardsen H, Touleimat N, Klajic J, Riis ML, Haakensen VD, Warnberg F, Naume B *et al*: Genome-wide DNA methylation profiles in progression to in situ and invasive carcinoma of the breast with impact on gene transcription and prognosis. *Genome biology* 2014, 15(8):435.
17. Stirzaker C, Zotenko E, Song JZ, Qu W, Nair SS, Locke WJ, Stone A, Armstong NJ, Robinson MD, Dobrovic A *et al*: Methylome sequencing in triple-negative breast cancer reveals distinct methylation clusters with prognostic value. *Nature communications* 2015, 6:5899.

18. Jin W, Li QZ, Zuo YC, Cao YN, Zhang LQ, Hou R, Su WX: Relationship Between DNA Methylation in Key Region and the Differential Expressions of Genes in Human Breast Tumor Tissue. *DNA and cell biology* 2019, 38(1):49-62.
19. Song B, Wang L, Zhang Y, Li N, Dai H, Xu H, Cai H, Yan J: Combined Detection of HER2, Ki67, and GSTP1 Genes on the Diagnosis and Prognosis of Breast Cancer. *Cancer biotherapy & radiopharmaceuticals* 2018.
20. Chen N, Zhao G, Yan X, Lv Z, Yin H, Zhang S, Song W, Li X, Li L, Du Z *et al*: A novel FLI1 exonic circular RNA promotes metastasis in breast cancer by coordinately regulating TET1 and DNMT1. *Genome biology* 2018, 19(1):218.
21. Dasgupta H, Islam S, Alam N, Roy A, Roychoudhury S, Panda CK: Hypomethylation of mismatch repair genes MLH1 and MSH2 is associated with chemotolerance of breast carcinoma: Clinical significance. *Journal of surgical oncology* 2019, 119(1):88-100.
22. Jahangiri R, Jamialahmadi K, Gharib M, Emami Razavi A, Mosaffa F: Expression and clinicopathological significance of DNA methyltransferase 1, 3A and 3B in tamoxifen-treated breast cancer patients. *Gene* 2019, 685:24-31.
23. Shilpi A, Bi Y, Jung S, Patra SK, Davuluri RV: Identification of Genetic and Epigenetic Variants Associated with Breast Cancer Prognosis by Integrative Bioinformatics Analysis. *Cancer informatics* 2017, 16:1-13.
24. Dietz S, Lifshitz A, Kazdal D, Harms A, Endris V, Winter H, Stenzinger A, Warth A, Sill M, Tanay A *et al*: Global DNA methylation reflects spatial heterogeneity and molecular evolution of lung adenocarcinomas. *International journal of cancer* 2019, 144(5):1061-1072.
25. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *et al*: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 2005, 102(43):15545-15550.
26. Wang J, Yu XF, N OU, Luo QL, Zhao SY, Guan XF, Chen T, Li JX: Multi-platform analysis of methylation-regulated genes in human lung adenocarcinoma. *Journal of toxicology and environmental health Part A* 2019:1-9.
27. Merino-Azpitarte M, Lozano E, Perugorria MJ, Esparza-Baquer A, Erice O, Santos-Laso A, O'Rourke CJ, Andersen JB, Jimenez-Aguero R, Lacasta A *et al*: SOX17 regulates cholangiocyte differentiation and acts as a tumor suppressor in cholangiocarcinoma. *Journal of hepatology* 2017, 67(1):72-83.
28. Chang WL, Lai WW, Kuo IY, Lin CY, Lu PJ, Sheu BS, Wang YC: A six-CpG panel with DNA methylation biomarkers predicting treatment response of chemoradiation in esophageal squamous cell carcinoma. *Journal of gastroenterology* 2017, 52(6):705-714.
29. Galamb O, Kalmar A, Peterfia B, Csabai I, Bodor A, Ribli D, Krenacs T, Patai AV, Wichmann B, Bartak BK *et al*: Aberrant DNA methylation of WNT pathway genes in the development and progression of CIMP-negative colorectal cancer. *Epigenetics* 2016, 11(8):588-602.
30. Balgkouranidou I, Chimonidou M, Milaki G, Tsaroucha E, Kakolyris S, Georgoulias V, Lianidou E: SOX17 promoter methylation in plasma circulating tumor DNA of patients with non-small cell lung cancer. *Clinical chemistry and laboratory medicine* 2016, 54(8):1385-1393.
31. Yao L, Shen H, Laird PW, Farnham PJ, Berman BP: Inferring regulatory element landscapes and transcription factor networks from cancer methylomes. *Genome biology* 2015, 16:105.
32. Fu D, Ren C, Tan H, Wei J, Zhu Y, He C, Shao W, Zhang J: Sox17 promoter methylation in plasma DNA is associated with poor survival and can be used as a prognostic factor in breast cancer. *Medicine* 2015, 94(11):e637.

33. Chimonidou M, Strati A, Malamos N, Georgoulias V, Lianidou ES: SOX17 promoter methylation in circulating tumor cells and matched cell-free DNA isolated from plasma of patients with breast cancer. *Clinical chemistry* 2013, 59(1):270-279.
34. Chimonidou M, Strati A, Tzitzira A, Sotiropoulou G, Malamos N, Georgoulias V, Lianidou ES: DNA methylation of tumor suppressor and metastasis suppressor genes in circulating tumor cells. *Clinical chemistry* 2011, 57(8):1169-1177.
35. Lim DH, Kim WS, Kim SJ, Yoo HY, Ko YH: Microarray Gene-expression Profiling Analysis Comparing PCNSL and Non-CNS Diffuse Large B-Cell Lymphoma. *Anticancer research* 2015, 35(6):3333-3340.
36. Lin Q, Ling YB, Chen JW, Zhou CR, Chen J, Li X, Huang MS: Circular RNA circCDK13 suppresses cell proliferation, migration and invasion by modulating the JAK/STAT and PI3K/AKT pathways in liver cancer. *International journal of oncology* 2018, 53(1):246-256.
37. Shang AQ, Wu J, Bi F, Zhang YJ, Xu LR, Li LL, Chen FF, Wang WW, Zhu JJ, Liu YY: Relationship between HER2 and JAK/STAT-SOCS3 signaling pathway and clinicopathological features and prognosis of ovarian cancer. *Cancer biology & therapy* 2017, 18(5):314-322.
38. Xu S, Kong D, Chen Q, Ping Y, Pang D: Oncogenic long noncoding RNA landscape in breast cancer. *Molecular cancer* 2017, 16(1):129.
39. Chen M, Pockaj B, Andreozzi M, Barrett MT, Krishna S, Eaton S, Niu R, Anderson KS: JAK2 and PD-L1 Amplification Enhance the Dynamic Expression of PD-L1 in Triple-negative Breast Cancer. *Clinical breast cancer* 2018, 18(5):e1205-e1215.
40. Khanna P, Lee JS, Sereemaspun A, Lee H, Baeg GH: GRAMD1B regulates cell migration in breast cancer cells through JAK/STAT and Akt signaling. *Scientific reports* 2018, 8(1):9511.
41. Radler PD, Wehde BL, Wagner KU: Crosstalk between STAT5 activation and PI3K/AKT functions in normal and transformed mammary epithelial cells. *Molecular and cellular endocrinology* 2017, 451:31-39.
42. Banerjee A, Jothimani G, Prasad SV, Marotta F, Pathak S: Targeting Wnt Signaling through Small molecules in Governing Stem Cell Fate and Diseases. *Endocrine, metabolic & immune disorders drug targets* 2019.
43. Lu ML, Zhang Y, Li J, Fu Y, Li WH, Zhao GF, Li XH, Wei L, Liu GB, Huang H: MicroRNA-124 inhibits colorectal cancer cell proliferation and suppresses tumor growth by interacting with PLCB1 and regulating Wnt/beta-catenin signaling pathway. *European review for medical and pharmacological sciences* 2019, 23(1):121-136.
44. Saha SK, Yin Y, Chae HS, Cho SG: Opposing Regulation of Cancer Properties via KRT19-Mediated Differential Modulation of Wnt/beta-Catenin/Notch Signaling in Breast and Colon Cancers. *Cancers* 2019, 11(1).
45. Wang F, Zhu W, Yang R, Xie W, Wang D: LncRNA ZEB2-AS1 contributes to the tumorigenesis of gastric cancer via activating the Wnt/beta-catenin pathway. *Molecular and cellular biochemistry* 2019.
46. Yang F, Xiao Z, Zhang S: Knockdown of miR-194-5p inhibits cell proliferation, migration and invasion in breast cancer by regulating the Wnt/beta-catenin signaling pathway. *International journal of molecular medicine* 2018, 42(6):3355-3363.
47. Farooqi AA, de la Roche M, Djamgoz MBA, Siddik ZH: Overview of the oncogenic signaling pathways in colorectal cancer: Mechanistic insights. *Seminars in cancer biology* 2019.
48. Gampa G, Kim M, Mohammad AS, Parrish KE, Mladek AC, Sarkaria JN, Elmquist WF: Brain distribution and active efflux of three panRAF inhibitors: considerations in the treatment of melanoma brain metastases. *The Journal of pharmacology and experimental therapeutics* 2019.
49. Abedini Bakhshmand E, Soltani BM: Regulatory effect of hsa-miR-5590-3P on TGFbeta signaling through targeting of TGFbeta-R1, TGFbeta-R2, SMAD3 and SMAD4 transcripts. *Biological chemistry* 2018.

50. Zhao Z, Hao D, Wang L, Li J, Meng Y, Li P, Wang Y, Zhang C, Zhou H, Gardner K *et al*: CtBP promotes metastasis of breast cancer through repressing cholesterol and activating TGF-beta signaling. *Oncogene* 2018.

Figures

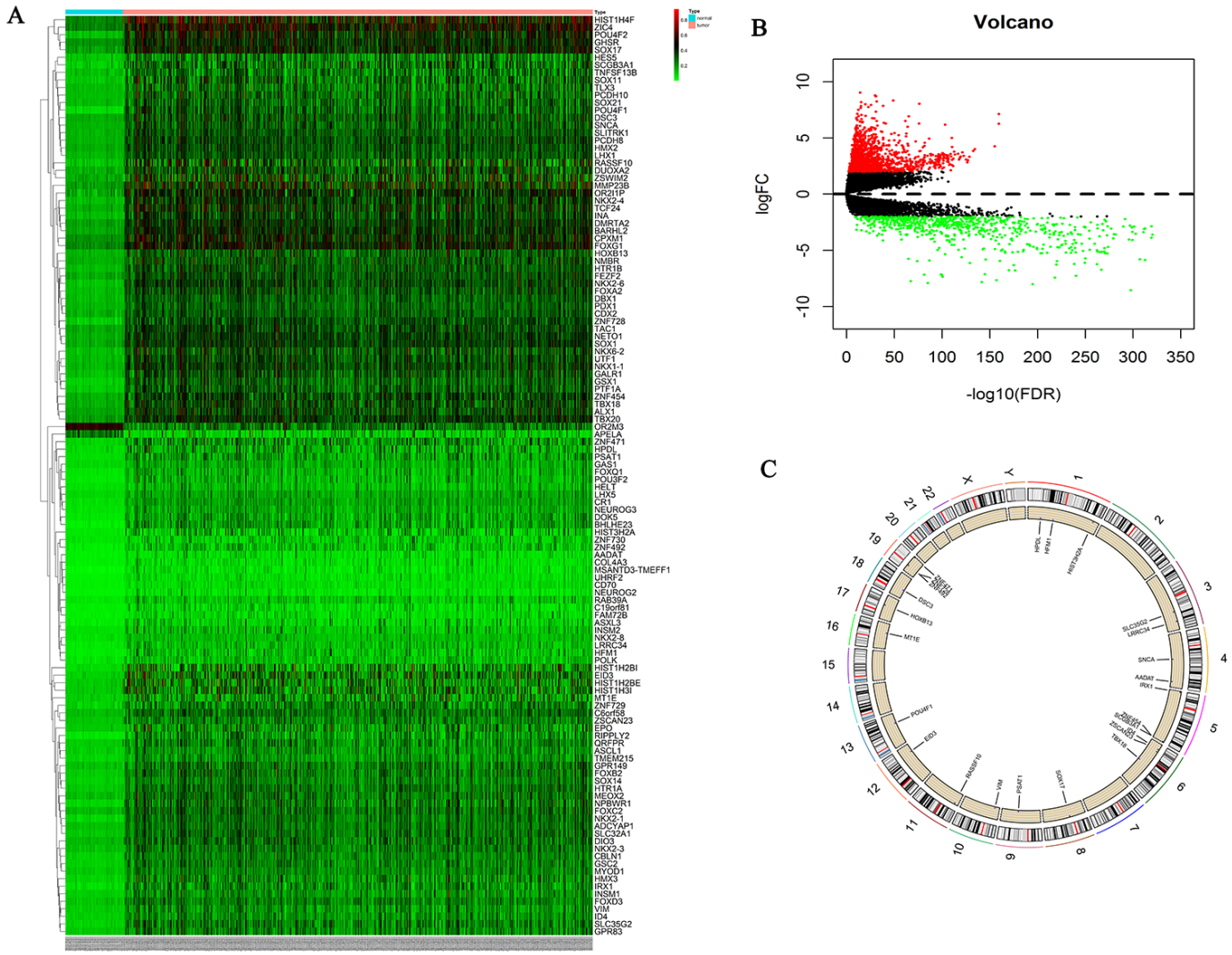

Figure 1

The difference analysis results obtained a total of 122 protein-coding genes

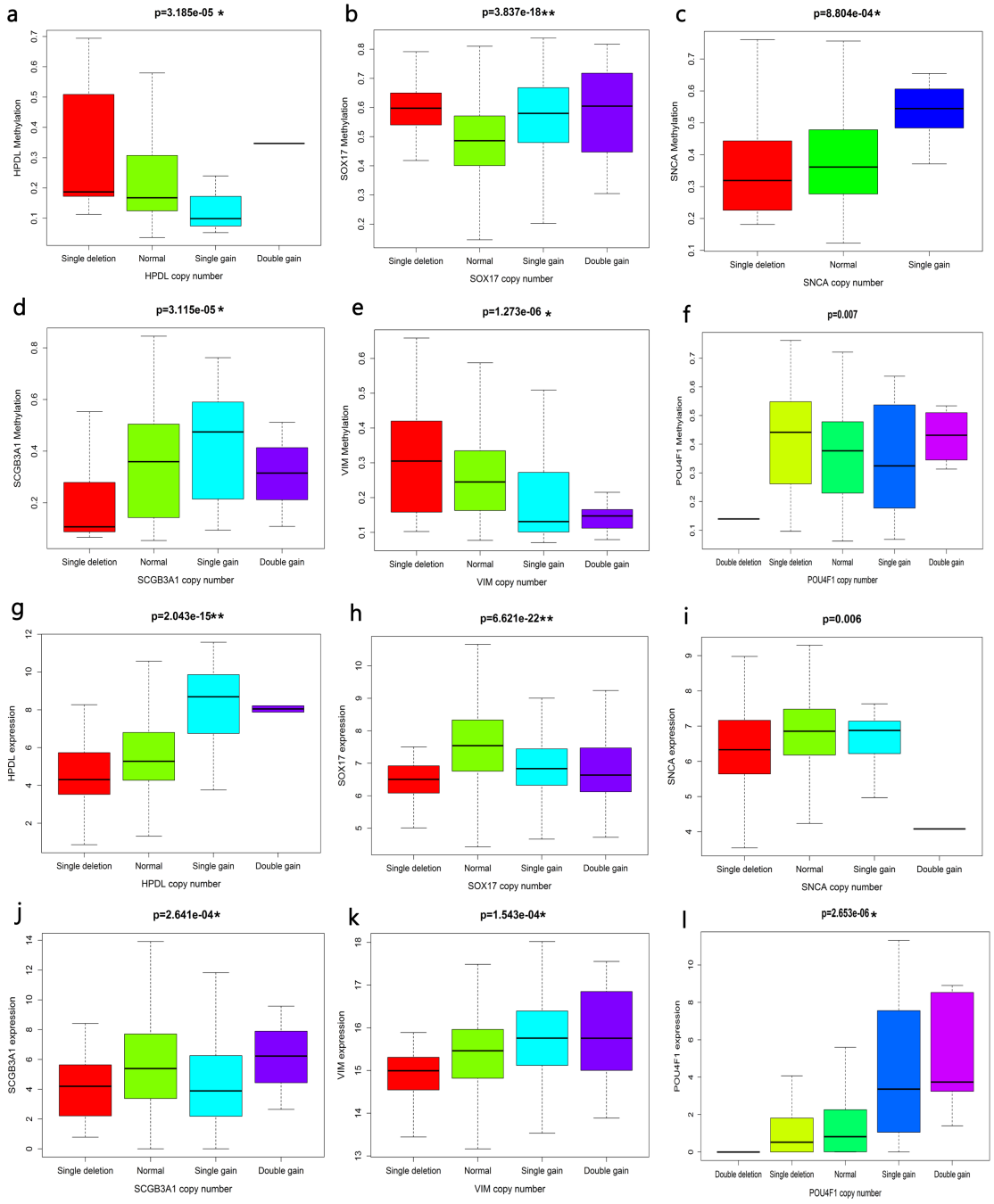

Figure 2

Six common genes.

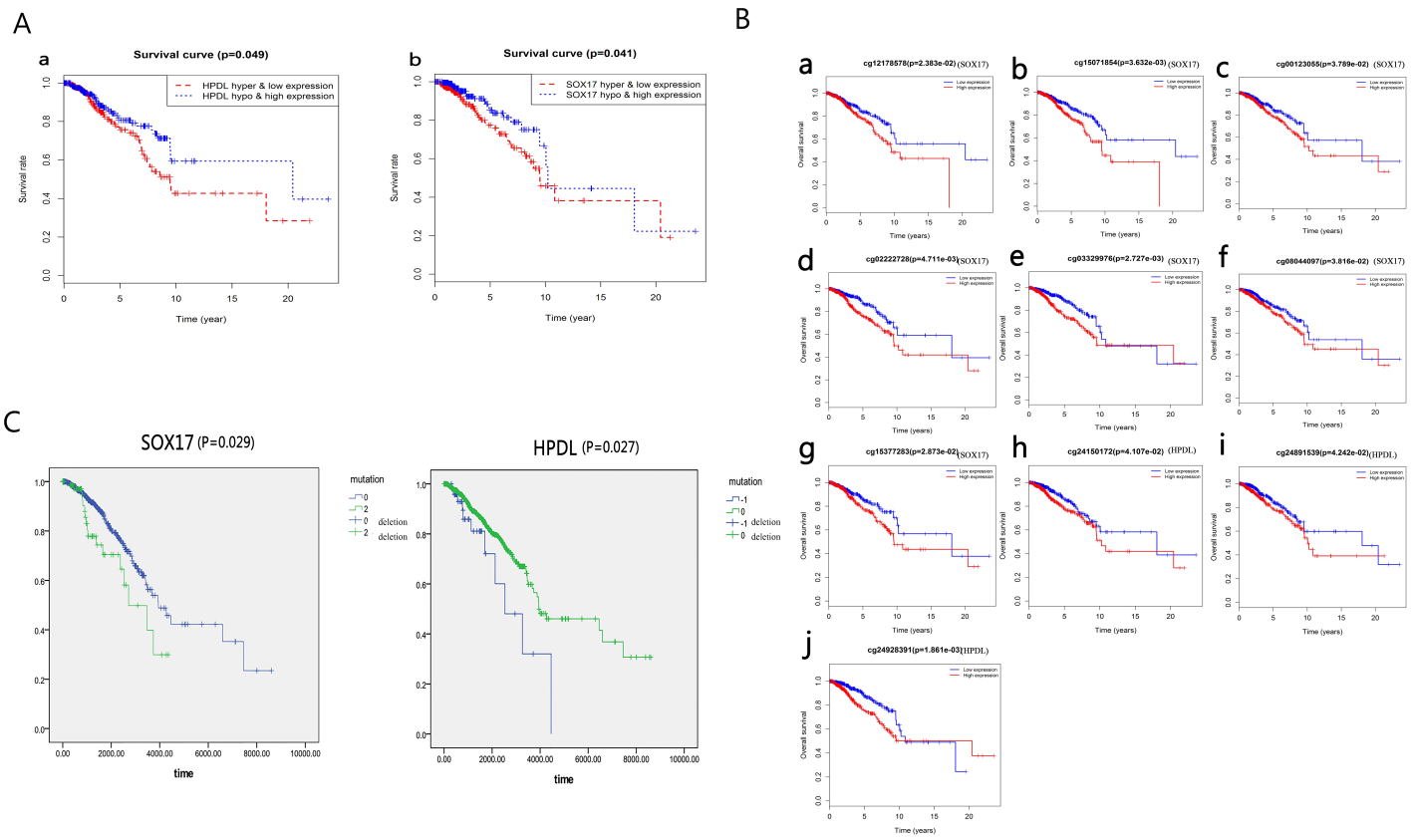

Figure 3

High-methylation low-expression of HPDL and SOX17 showed poor prognosis

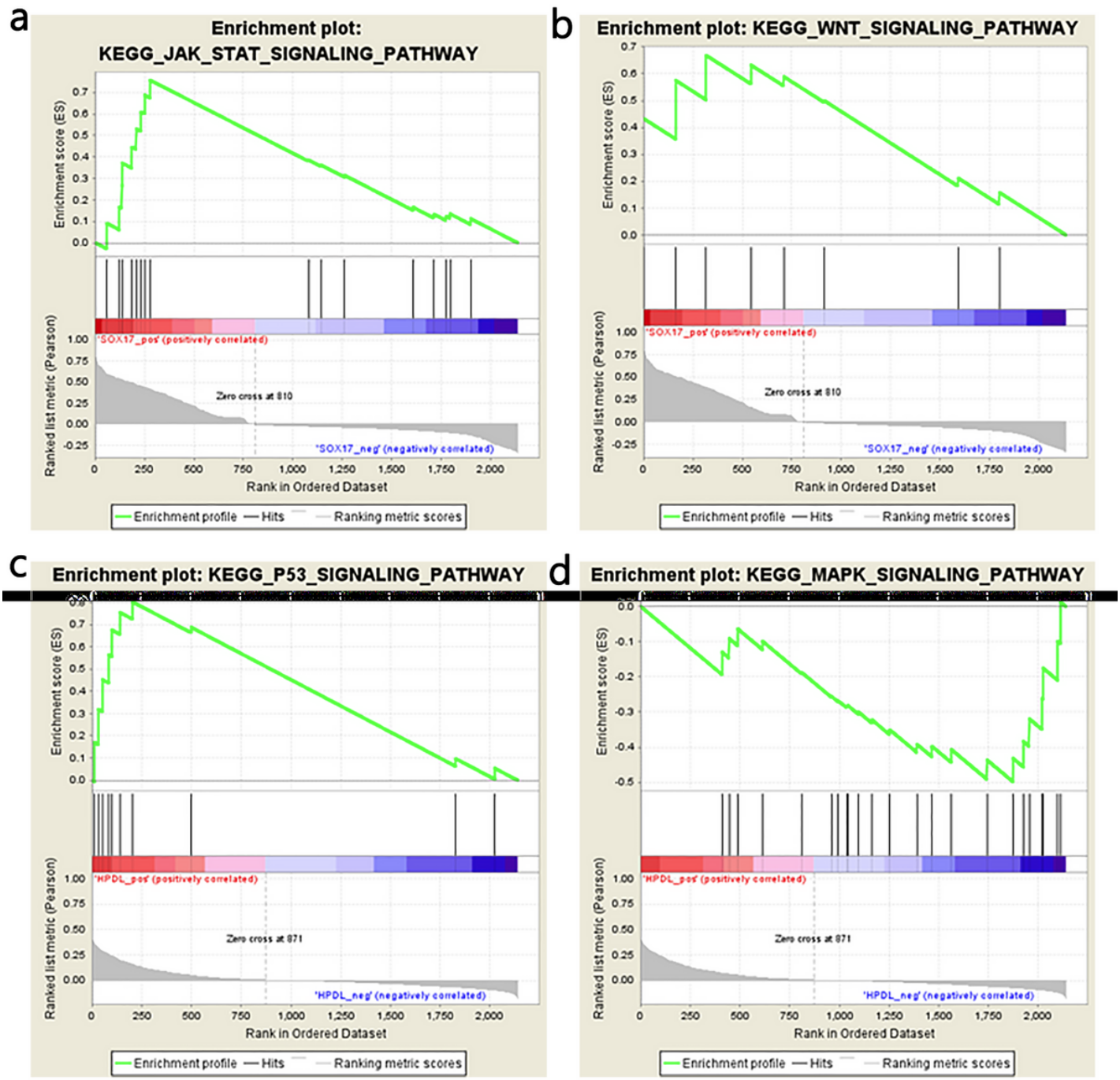

Figure 4

SOX17 mainly affects JAK STAT signaling pathway, WNT signaling pathway, and so on

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial1.txt
- Supplementarymaterial2B.tif
- Supplementarymaterial2A.tif