Genome-Wide DNA Methylome and Whole-Transcriptome Landscapes of Spontaneous Intraductal Papilloma in Tree Shrews (*Tupaia Belangeri*)

Chengxiu Liu

Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Medical Biology

Yuanyuan Han

Institute of Medical Biology, Chinese Academy of Medical Sciences

Pinfen Tong

Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Medical Biology

Dexuan Kuang

Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Medical Biology

Na Li

Chinese Academy of Medical Sciences Institute of Basic Medical Sciences Department of Epidemiology and Biostatistics

Caixia Lu

Institute of Medical Biology, Chinese Academy of Medical Sciences

Xiaomei Sun

Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Medical Biology

Wenguang Wang

Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Medical Biology

Jiejie Dai (✉ djj@imbcams.com.cn)

Institute of Medical Biology <https://orcid.org/0000-0002-4242-9089>

Research article

Keywords: Tree shrew, Spontaneous breast intraductal papilloma, Methylation, Transcriptome

Posted Date: December 8th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-121950/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

Breast intraductal papilloma (IP) is mainly caused by the abnormal proliferation of ductal epithelial cells. Tree shrews are a potential animal model for studying breast tumours. However, little is known about the transcriptome and DNA methylome landscapes of breast IP in tree shrews. In this work, we performed whole-genome DNA methylation and transcriptome analyses of breast IPs and normal mammary glands in tree shrews.

Results

DNA methylation profiles with a single-base resolution were generated via whole-genome bisulphate sequencing (WGBS) in both the IP and Control groups of tree shrews. This provided a genome-wide perspective regarding the epigenetic regulation of protein-coding genes in breast IP in tree shrews. The methylation levels at CG sites were considerably higher than those at CHG and CHH sites, and methylation levels were highest in gene body regions. We identified 3486, 82 and 361 differentially methylated regions (DMRs) in CG, CHG, and CHH contexts, respectively, and 701 differentially methylated genes (DMGs) were found. Furthermore, transcriptomic analysis identified 62 differentially expressed genes (DEGs), 50 long noncoding RNAs (lncRNAs), and 32 circular RNAs (circRNAs) in IPs compared with normal mammary glands. Correlation analysis between the DNA methylation and transcriptome data showed that 25 DMGs were also DEGs, among which the expression levels of 9 genes were negatively correlated with methylation levels in gene body regions. Importantly, integrated analysis identified three genes, PDZK1, ATP2B4 and LCP1, that could be used as candidates for further studying breast IP in tree shrews.

Conclusions

Overall, this research provides the comprehensive landscape of the transcriptome and DNA methylome of spontaneous IP in tree shrews and highlights candidate genes for eliciting tumorigenesis. These results contribute to the application of tree shrews as an animal model of breast tumours.

Background

Intraductal papilloma (IP) is a benign tumour found within breast ducts that accounts for approximately 10% of benign breast lesions [1]. The abnormal proliferation of ductal epithelial cells causes such growth, which occurs most commonly in women between 35–55 years of age [2]. Hormones, fertility, and diet are predisposing risk factors that may lead to the development of IPs [3]. Because IP is related to atypia, ductal carcinoma in situ (DCIS) and carcinoma, it is classified as a high-risk precursor lesion with a 6.3% risk of being malignant [3] and may be upgraded to atypical ductal hyperplasia or DCIS upon surgical excision [4]. However, the mechanism of breast neoplasm is not fully understood, and multidimensional molecular data from IP patients has not been fully integrated in studies on this topic.

The results of DNA sequence research confirm that tree shrews are close relatives of primates [5]. Tree shrews have become an increasingly popular experimental animal model for a variety of human tumour diseases, including lung cancer [6], hepatocellular carcinoma [7] and glioblastoma [8]. The genome sequencing of Chinese tree shrews was accomplished in 2013 and has provided a useful resource for functional genomic studies [9]. A database based on the genome sequence data of tree shrews has been established [10]. Most importantly, in terms of morphology and structure, the mammary glands of tree shrews are similar to human glands [11]. Based on the above advantages, tree shrews are ideal experimental animals for studying the pathogenesis of mammary tumours. However, there are few studies on gene expression patterns and the underlying function of DNA methylation in the tumorigenesis of spontaneous IP in tree shrews as a novel breast tumour animal model.

DNA methylation is one of the epigenetic changes that has been shown to play an important role in the pretranscription regulation and inhibition of gene expression in many mammalian genomes. The mapping of genome-wide DNA methylation is of great importance for understanding tumorigenesis [12]. This modification participates in many cancers, including thyroid cancer [13], non-small cell lung cancer [14], gastric cancer [15], etc. DNA methylation is involved in the development and progression of breast cancer [16]. Likewise, limited evidence has suggested that the aberrant methylation of cytosine residues is involved in the development of IPs [17]. Therefore, it is helpful to understand the tumorigenesis of papilloma from the perspective of epigenetic regulation by depicting the methylation profile and identifying differentially methylated genes (DMGs).

It is generally believed that the abnormal reprogramming of the whole transcriptome, including genes, long noncoding RNAs (lncRNAs) and circular RNAs (circRNAs), is a crucial process in tumour occurrence and progression. Recently, RNA sequencing (RNA-seq) studies on breast cancer [18] have been conducted to provide a deeper understanding of the mechanisms involved, and research on the potential underlying molecular mechanisms that influence the occurrence and development of breast cancer has been performed in murine mammary tumour models [19]. By analysing methylome and transcriptome variations related to the survival status of breast cancer, we can further understand the basic biological process of breast cancer based on the genetic aetiology [20]. Moreover, the results of the analysis of DNA methylation and gene expression have demonstrated that the methylation level of CpGs in breast cancer tissues is significantly higher than that in adjacent normal tissues. Additionally, large numbers of CpGs show a significantly higher methylation level than that found in nearby normal tissues, which is negatively correlated with gene expression [21]. Thus, combined with methylation data and gene expression profile data, we can better analyse the regulatory function of methylation to solve existing problems.

However, compared to studies in malignant breast cancer, studies on DNA methylation and the transcriptome in IPs are lagging behind. Therefore, we carried out an integrated analysis of genome-wide DNA methylation levels and the whole transcriptome in breast IP in tree shrews in the present work. Our study provides new insights into IP in tree shrews, highlights candidate tumorigenesis-eliciting genes, and will contribute to the application of tree shrews as breast tumour animal models.

Results

Diagnosis and pathology identification of tree shrew spontaneous breast IPs

In the basic pathology of tree shrew breast tumours, the ductal epithelium of the breast showed papillary hyperplasia, the nipple varied in size, and the cells showed no atypia, similar to human pathology. The tumours are expected to be benign (Fig. 1A). In contrast, the normal mammary glands showed a normal structure of the breast acini and ducts, and there was no degeneration, necrosis, or inflammatory cell infiltration (Fig. 1B). As shown in Fig. 1, we confirmed that 3 of the spontaneous mammary tumours collected from females in the closed colony of tree shrews were breast IPs. The maximum size of the tree shrew spontaneous breast IPs was 5 cm x 4 cm. As shown in Table 1, the selected tree shrews were divided into two groups. The IP group (n = 3) consisted of tree shrews with IPs (IP-1, IP-2 and IP-3), whereas the Control group (n = 3) consisted of tree shrews with healthy mammary gland tissues (Control-1, Control-2 and Control-3). These results indicated that the selected IP tree shrews and normal tree shrews were appropriate for subsequent analyses.

Table 1
Basic information of the six tree shrew tissue samples.

Group	Sample ID	Age	Tumor location	Tumor size (mm ³)
IP	IP-1	5	Subcutaneous of lower abdomen	52333
	IP-2	4	Subcutaneous of left chest	12821
	IP-3	6.5	On the left side of the abdomen	39577
Control	Control-1	5	—	—
	Control-2	5	—	—
	Control-3	5	—	—

Genome-wide DNA methylation profiling of breast IPs and normal mammary gland tissues of tree shrews

To investigate methylation patterns during tree shrew IP development, we analysed genome-wide DNA methylation (DNAm) levels in the IP group and normal mammary gland tissue by whole-genome bisulphate sequencing (WGBS) with > 99% conversion efficiency. The genome of the normal mammary gland (Control) presented ~ 4.39% methylated cytosines (mCs), and the IP sample presented ~ 4.41% mCs among the total sequenced C sites, reflecting the degree of genome methylation. The methylation levels of CG, CHH, and CHG (where H is A, C, or T) sites were distinct. We detected genome-wide mC levels of 88.08 ± 1.76% for CG, 2.52 ± 0.32% for CHG, and 9.40 ± 1.45% for CHH in the Control group and 90.19 ± 0.54% for CG, 2.07 ± 0.41% for CHG, and 7.74 ± 0.51% for CHH in the IP group. Thus, the proportions of these contexts were analogous between groups (Fig. 2A).

Epigenetic variation among DNA sequences, particularly CpG DNA methylation, is an important type of variation that modulates gene expression under physiological and pathological conditions. Our results showed that CG methylation levels were higher while CHG and CHH methylation levels were lower and that the number of mCG sites was increased in the IP group compared with the Control mammary gland tissues, while other types of sites did not differ significantly. To further compare the coding gene distribution as well as the methylation levels of diverse functional genomic elements between the two groups, we analysed the DNA methylation pattern in three distinct regions of transcriptional elements: the upstream 2 k region (2000 bp before the transcription start site (TSS) of the gene), the gene body region, and the downstream 2 k region (2000 bp after the gene termination site). The distribution characteristics of DNA methylation levels in distinct functional regions help to understand the characteristics of DNA methylation modifications in different regions at the whole-genome level.

The obtained DNA methylation profiles demonstrated that the methylation level in the CG context was higher than those in the CHG and CHH contexts. The DNA methylation level in the CG context was highest in the gene body region. DNA methylation was moderately high in the upstream 2K start site, decreased dramatically from the upstream 2 k region to the TSS, increased sharply from the TSS to the gene body region, maintained the highest level in the gene body region and then decreased slightly in the downstream 2K region (Fig. 2B).

We identified 3,486 differentially methylated CG regions, 82 CHG regions, and 361 CHH regions. In the CG context, 3,486 differentially methylated regions (DMRs), located in 701 genes, were identified between the IP and Control groups ($Q < 0.05$), among which 705 showed increased methylation and 2,781 showed decreased methylation in IP tissues compared with the levels in Control tissues (Additional file 1: Table S1). All DMRs were used for the comparison of differences, as shown in Fig. 3A. We analysed CG methylation sites in tree shrews in the IP and Control groups using hierarchical clustering. The results revealed separation between the two groups.

CG site methylation levels for the total methylated sites and DMGs in IP tumours (69.47%) were decreased compared with those in mammary glands (74.18%). Among the genes containing DMRs, 607 genes were located within the gene body, including 259 upregulated genes and 357 downregulated genes. A total of 58 genes were located in the upstream 2K region, whereas 55 were located in the downstream 2K region (Fig. 3B) (Additional file 1: Table S1).

To further elucidate the biological functions of DMGs, we performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. Based on GO term analysis, the putative target genes of KLF5 were associated with terms such as developmental process ($P_{\text{adjust}} = 0.0056$) and single-organism developmental process ($P_{\text{adjust}} = 0.0103$) in the biological process (BP) category; binding ($P = 0.0063$) in the molecular function (MF) category; and membrane ($P = 0.0045$) in the cellular component (CC) category (Fig. 3C) (Additional file 1: Table S1). KEGG pathway analysis indicated that 15 DMGs were associated with the oxytocin signalling pathway ($Q = 0.0027$), while 10 DMGs were associated with the oestrogen signalling pathway ($Q = 0.0113$).

(Fig. 3D) (Additional file 1: Table S1). Taken together, the gene annotation results revealed that the majority of differentially genes showing increased methylation and decreased methylation were mapped to the gene body region, suggesting that DMGs play crucial roles in IP.

Differentially expressed genes (DEGs) between the IP group and normal mammary gland tissue

To systematically describe the transcriptome landscape of the IP group and normal mammary gland tissues, whole-transcriptome sequencing was performed. After removing low-quality reads from each library, the clean reads were combined and aligned with the *Tupaia chinensis* genome (TupChi_1.0), resulting in the identification of 10,051 known mRNAs, 25,481 new mRNAs, 1,022 known lncRNAs, 1,617 new lncRNAs, and 10,602 new circRNAs in IPs (IP group) and normal mammary glands (Control group) in total. Furthermore, we compared the transcriptomic landscapes of the normal mammary gland and IP tumour tissues of tree shrews. We identified 39 upregulated DEGs and 23 downregulated DEGs in IP tumours compared with normal mammary glands ($FDR < 0.05$ & $|\log_2FC| > 1$) (Fig. 4A). Among these DEGs, the expression of ST14, PSMF1, and TNFSF1 was more than 15-fold higher in the IP group. In normal mammary gland tissue, the levels of FAM192A and Psmc5 were 15- and 7-fold higher, respectively. More detailed information is listed in Additional file 1: Table S2.

Among these DEGs, GO analysis indicated that TNFSF11 participates in the tumour necrosis factor-mediated signalling pathway and the cellular response to tumour necrosis factor ($P = 0.0042$); GATA3, EPHA2, and PTPRG are involved in the regulation of epithelial cell migration ($P = 0.0042$); STAG2, TEX14, PPP1R9B, and TPR are involved in the regulation of cell cycle processes ($P = 0.0087$); and TNFSF11, and EPHA2 are involved in epithelial cell proliferation ($P = 0.0091$) (Additional file 1: Table S2). ACSL1 and APOA5 participate in the PPAR signalling pathway ($P = 0.0367$) according to KEGG analysis (Additional file 1: Table S2). Furthermore, we utilized gene set enrichment analysis (GSEA) to analyse groups of functionally relevant genes in IPs compared to normal mammary glands. A total of 107 GO terms and 66 pathways ($FDR < 0.05$) were indicated to be significantly enriched in dysregulated genes in IPs (Additional file 1: Table S2). The upregulation of cell cycle phase- and S phase-related terms in the biological process category and mitotic sister chromatid segregation and sister chromatid segregation in the cellular component organization or biogenesis category was observed under the BP category in the IP group (Additional file 1: Table S2). Among the DEGs identified in the IP group, the upregulated gene sets were involved not only in replication and repair, including DNA replication, mismatch repair and nucleotide excision repair, but also in cell growth and death, including the cell cycle; the downregulated gene sets were involved in the endocrine system, including the renin secretion and PPAR signalling pathways (Additional file 1: Table S2). Overall, many of the DEGs identified in IPs were found to be involved in tumorigenesis, as demonstrated by the enrichment analysis of GO terms and pathways.

The lncRNA expression profile is distinct between IPs and normal mammary gland tissue in tree shrew

Compared with the results in normal mammary gland tissue samples, 50 differentially expressed lncRNAs (DElncRNAs), including 39 upregulated and 11 downregulated lncRNAs, were identified in the IP tissue samples (Fig. 5A). The most significantly upregulated lncRNAs were TCONS_00002926, TCONS_00077528, XR_001369927.1, TCONS_00037489, and XR_333956.1, whereas TCONS_00016941, XR_334171.1, TCONS_00135842, TCONS_00077456 and TCONS_00094673 were downregulated lncRNAs identified in the IP group (Additional file 1: Table S3). To reveal the function of the lncRNAs, we predicted the complementary correlation of antisense lncRNAs and mRNAs. All antisense target genes of the DElncRNAs were subjected to GO and KEGG pathway analysis to determine their functions. A variety of relevant GO terms in the BP category were observed (Additional file 1: Table S3), such as blood vessel endothelial cell migration ($P = 3.99E-05$), epithelial cell migration ($P = 0.0012$), epithelium migration ($P = 0.0012$), and negative regulation of apoptotic process ($P = 0.0016$). KEGG pathway analysis revealed that the Jak-STAT signalling pathway ($P = 0.0087$), PPAR signalling pathway ($P = 0.0233$), and oestrogen signalling pathway ($P = 0.0292$) were associated with DElncRNAs (Additional file 1: Table S3).

The second function of lncRNAs, when located less than 10 kb upstream/downstream from a gene, is to act as cis-regulators of their neighbouring genes on the same strand. A large number of enriched GO terms were observed in the BP category, including branching morphogenesis of an epithelial tube ($P = 0.0001$), serine phosphorylation of STAT protein ($P = 0.0002$), negative regulation of cell growth ($P = 0.0017$), and execution phase of apoptosis ($P = 0.0019$) (Additional file 1: Table S3). KEGG pathway analysis revealed that DElncRNAs were associated with the pathways of protein processing in the endoplasmic reticulum ($P = 0.0017$), glyoxylate and dicarboxylate metabolism ($P = 0.0018$), antigen processing and presentation ($P = 0.0250$), and the NOD-like receptor signalling pathway ($P = 0.0483$) (Additional file 1: Table S3).

The third function of lncRNAs is the trans-regulation of non-adjointing co-expressed genes. We analysed the correlation of expression between lncRNAs and mRNAs to reveal the target genes of lncRNAs and performed GO function and KEGG pathway enrichment analyses of protein-coding genes with an absolute correlation greater than 0.9 (Additional file 1: Table S3). The enriched GO functions of the DElncRNA trans-regulated co-expressed genes are listed in Table S3 and included mesenchymal stem cell differentiation ($P = 0.0032$), extrinsic apoptotic signalling pathway ($P = 0.0070$), recombinational repair ($P = 0.0086$), and the execution phase of apoptosis ($P = 0.0125$) in the BP category. The DElncRNA trans-regulated co-expressed genes were significantly enriched in KEGG pathways including homologous recombination ($P = 0.0004$), histidine metabolism ($P = 0.0022$), and the renin-angiotensin system ($P = 0.0244$) (Additional file 1: Table S3). These dysregulated lncRNAs might be involved in IP tumorigenesis in tree shrews by regulating genes and signalling pathways implicated in tumorigenesis.

The circRNA expression profile is distinct between IP and normal mammary gland tissues of tree shrews

As shown in Fig. 6A, we identified 32 differentially expressed circRNAs (DEcircRNAs), including 25 upregulated circRNAs and 7 downregulated circRNAs, between IPs and normal mammary glands ($P < 0.05$ & $|\log_2FC| > 1$). In the heat map of DEcircRNAs, IP tissues could be separated from normal mammary gland tissues by the DEcircRNAs. The most significantly upregulated circRNAs were novel_circ_004184, novel_circ_001608, novel_circ_007270, novel_circ_004893, and novel_circ_004886, while novel_circ_007552, novel_circ_007844, novel_circ_002826, novel_circ_005946 and novel_circ_009457 were identified as downregulated circRNAs in the IP group (Additional file 1: Table S4).

The gene of origin of a circRNA is its parental gene. We carried out a functional enrichment analysis of parental genes to investigate the putative functions of differentially expressed circRNAs. GO analyses revealed that the parental genes of the dysregulated circRNAs were enriched in the terms branch elongation of an epithelium ($P = 0.0161$), negative regulation of cell proliferation ($P = 0.0173$), positive regulation of epithelial cell proliferation ($P = 0.0233$), vasculogenesis ($P = 0.0276$), and gland morphogenesis ($P = 0.0290$) in the BP category. The PI3K-Akt signalling pathway ($P = 0.0139$) and Ras signalling pathway ($P = 0.0509$) were identified in the KEGG enrichment analysis (Fig. 6B, 6C, Additional file 1: Table S4). Taken together, these findings suggest that these DEcircRNAs that influence gene expression may affect IP tumour development.

Integration analysis of the methylome and transcriptome

To examine whether differences in DNA methylation between the IP and healthy mammary gland tissues of tree shrews could be the basis of the observed gene expression differences, we analysed the correlation between the gene expression and DNA methylation data in the IP group against that in the normal mammary gland group, and the results indicated a considerable regulatory effect of DNA methylation on the modulation of gene expression. The methylation levels in the upstream 2K (Pearson's $R = -0.0383$ for IP group; and Pearson's $R = -0.0192$ for Control group), gene body (Pearson's $R = -0.0922$ for IP group; and Pearson's $R = -0.0737$ for Control group) and downstream 2K regions in the Control group (Pearson's $R = -0.0013$ for Control group) were negatively correlated with expression levels, but those in the downstream 2K region were positively correlated in the IP group (Pearson's $R = 0.0030$ for IP group) (Fig. 7A, 7B).

Additionally, DMGs and DEGs were compared through integrated methylomic and transcriptomic analysis, which showed 25 differentially methylated and expressed genes according to both RNA-seq ($P < 0.05$) and WGBS ($Q < 0.05$) (Fig. 7B and Additional file 1: Table S5). Among these genes, the gene numbers exhibiting DMRs in the upstream 2K, gene body and downstream 2K regions were 1, 23 and 1, respectively. The GO analysis of DEGs and DMGs showed that 15 genes could show enrichment in 213 BP terms ($P < 0.05$) (Additional file 1: Table S5). Among these genes, ATP2B4 was involved in the positive regulation of transmembrane transport ($P = 0.0008$), calcium-mediated signalling ($P = 0.0022$), the negative regulation of catabolic processes ($P = 0.0031$), and the negative regulation of calcium-mediated signalling ($P = 0.0061$); PDZK1 was involved in the positive regulation of transmembrane transport ($P = 0.0014$), positive regulation of transport ($P = 0.0031$) and regulation of transmembrane transport ($P = 0.0126$); LCP1 was involved in the regulation of intracellular transport ($P = 0.0034$) and the regulation of cellular localization ($P = 0.0151$); and PDZK1 and LCP1 are involved in the regulation of transport ($P = 0.0135$). The KEGG analysis of DEGs and DMGs showed that 9 genes were enriched in 19 signalling pathways ($P < 0.05$) (Additional file 1: Table S5). Subsequent analysis identified 9 genes with an inverse relationship between their degree of DNA methylation and gene expression in gene body regions (Table 2), which were related to signal transduction and the endocrine system. Three differentially upregulated and downregulated genes (ADCY5, ATP2B4, and CREB5) were associated with signal transduction pathways, including the cAMP signalling pathway ($P = 3.13E-06$), cGMP-PKG signalling pathway ($P = 6.76E-05$), TNF signalling pathway ($P = 0.0118$), and AMPK signalling pathway ($P = 0.0131$). Furthermore, ADCY5 and CREB5 were involved in insulin secretion ($P = 0.0001$) and the oestrogen signalling pathway ($P = 0.0002$) in the endocrine system. Thus, we integrated the gene expression and DNA methylation maps and identified protein-coding genes with underlying changes related to DNA methylation in IPs; the resulting alterations probably induced the development of mammary tumours.

Table 2
Significant DEGs showing a negative correlation with DNA methylation involved in the development of intraductal papilloma.

Gene ID	start	end	DNA methylation			mRNA			Symbol		
			P value	Q value	Methylation difference	Control fpkm	IP fpkm	log2 (FC)	P value		Description
XM_006146258.2	69740001	69740200	1.09E-20	4.56E-18	-35.605	0.001	1.370	10.420	0.008	ADGRL3	latrophilin-3 isoform X6
XM_006146258.2	69467601	69467800	2.00E-16	4.18E-14	-33.942	0.001	1.370	10.420	0.008	ADGRL3	latrophilin-3 isoform X6
XM_006146258.2	69467501	69467700	1.16E-17	3.02E-15	-33.442	0.001	1.370	10.420	0.008	ADGRL3	latrophilin-3 isoform X6
XM_014582091.1	139578801	139579000	1.21E-09	5.35E-08	-27.965	20.140	95.463	2.245	0.021	LCP1	plastin-2
XM_014582091.1	139578901	139579100	1.21E-09	5.35E-08	-27.965	20.140	95.463	2.245	0.021	LCP1	plastin-2
XM_006170445.2	23207301	23207500	4.60E-13	4.82E-11	-26.329	0.001	8.303	13.019	0.044	UBE2V2	Ubiquitin-conjugating enzyme E2 va 2
XM_006169173.2	65590701	65590900	1.74E-14	2.49E-12	27.634	179.237	50.200	-1.836	0.035	DNASE1L3	deoxyribonucl gamma
XM_014583774.1	35152901	35153100	5.78E-06	7.69E-05	28.147	8.087	0.223	-5.178	0.037	CREB5	cyclic AMP-responsive element-binding protein 5 isoform X3
XM_006157144.2	118399201	118399400	2.95E-09	1.16E-07	29.807	11.083	0.883	-3.649	0.045	PDZK1	Na(+)/H(+) exchange regulatory cofactor NHE-
XM_006154597.1	40832101	40832300	2.93E-18	8.43E-16	33.982	4.490	0.170	-4.723	0.007	ADCY5	adenylate cyclase type 5
XM_006164776.2	106988901	106989100	4.92E-14	6.38E-12	39.173	7.483	0.993	-2.913	0.029	ADAMTSL1	ADAMTS-like protein 1
XM_006159177.2	28289501	28289700	1.30E-24	8.60E-22	45.829	4.847	0.030	-7.336	0.002	ATP2B4	plasma membrane calcium-transporting ATPase 4

Start/end: the gene position from start to end; Q value: adjusted p value; Control fpkm/IP fpkm: gene expression levels in the Control/IP group; log2 (FC): log₂ the fold change in gene expression. A positive value indicates upregulation, and a negative value indicates downregulation.

Discussion

Herein, we provide an expanded overview of DNA methylation levels and transcriptome characteristics in 3 tumour tissue samples of spontaneous breast IP and normal mammary gland tissues from tree shrews in which gene expression was analysed. Tree shrews are currently considered an animal model for studying mammary tumours, including both spontaneous and induced models. Kazuhiro Daino et al. used a microarray to obtain the DNA methylation and expression profiles of γ -ray-induced mammary carcinomas in rats [22]. We obtained the genomic DNA methylation profiles of normal mammary gland and IP tissues of tree shrews in this context. DNA methylation levels were downregulated in tree shrews with IP compared with healthy tree shrews. In addition, GATA-binding protein 3 (GATA3) was upregulated 12.2-fold in IP tissues compared with normal mammary gland tissues. However, no difference in DMRs was found. The different results between these two studies may be due to the differences in the experimental animals and types of mammary tumours examined.

Shao et al identified 17 Krüppel-like factors from Chinese tree shrews. KLF5 encodes a member of the zinc finger protein KLF subfamily that acts a transcriptional activator by binding to a specific recognition motif directly in the promoters of target genes to play roles in both promoting and suppressing cell proliferation. *Tupaia belangeri* (Tb) KLF5, similar to human Krüppel-like factor (hKLF) hKLF5, significantly promotes cell proliferation, playing a pro-proliferative and oncogenic role in breast cancer [23]. These findings suggested that tree shrews may serve as alternative animal models for breast cancer related to KLF5 [24]. In this study, KLF family members were not included among the DEGs, but DNA methylation analysis showed that KLF5 presented significant upregulation among the identified DMRs.

To obtain deeper insights into the molecular mechanisms of tumorigenesis, we carried out a correlation analysis of the RNA-seq and WGBS data. The results suggested an inverse correlation of PDZK1, ATP2B4 and LCP1 gene expression with DNA methylation between the IP group and Control groups. PDZ domain-containing 1 (PDZK1) encodes a PDZ domain-containing scaffolding protein [25]. Genome-wide association studies (GWASs) of a large cohort identified rs12405132 of PDZK1 as a new susceptibility locus for breast cancer; hence, PDZK1 is a potential interacting gene in breast cancer [26]. In primary breast

cancers, PDZK1 is an oestrogen-regulated gene that is overexpressed in ER-positive breast cancers [27]. PDZK1 has been as a marker of oestrogen-regulated gene (ERG) expression in examining the relationship between the menstrual cycle and oestrogen receptor-positive breast cancer [28]. Dunbier et al. further verified that the expression of PDZK1 was strongly related to plasma oestradiol (E2) levels in postmenopausal patients with primary ER-positive breast cancer [29]. PDZK1 exhibits epithelial expression with a primarily cytosolic subcellular localization, and PDZK1 expression is indirectly modulated by ER- α stimulation [30].

ATPase Plasma Membrane Ca^{2+} Transporting 4 (ATP2B4) encodes plasma membrane calcium ATPase isoform 4 (PMCA4b), which belongs to the P-type primary ion transport ATPase family. The PMCA4b (ATP2B4) protein is located primarily in the plasma membrane, is expressed in normal breast tissue, and plays an important role in the plasma membrane Ca^{2+} pump in the maintenance of mammary epithelial Ca^{2+} homeostasis [31]. The PMCA4 protein is present in the normal breast ductal epithelium; however, a variety of factors, including hormonal imbalances, epigenetic modifications and impaired protein trafficking, may lead to PMCA4b loss in breast cancer [32]. The same study showed that the regulation of Ca^{2+} signalling by increased PMCA4b expression may be conducive to the normal development of the breast epithelium. Consistent with the results of this previous study, the expression of ATP2B4 mRNA was downregulated in our IP group by 7.3-fold, and methylation was upregulated. In breast cancer treatment, the targeting of PMCA4 may enhance the effectiveness of breast cancer therapies that act through the promotion of cell death pathways [33]. The targeted regulation of PMCA4 functionality may give rise to novel therapeutic methods to attenuate or facilitate new vessel formation in breast cancer, which is associated with angiogenesis [34].

Lymphocyte cytosolic protein 1 (LCP1) is an L-plastin protein-coding gene that is a member of the actin-binding protein family, which is conserved during eukaryote evolution and is expressed in the majority of tissues of higher eukaryotes. LCP1 plays a critical role in the activation of T-cells, associated with NF- κ B signalling, calcium ion binding and actin binding. In addition, the expression of L-plastin is induced concomitant with tumorigenesis in solid tissues. A negative effect of LCP1 on breast cancer progression has been proven, and the inhibition of LCP1 results in breast cancer cell migration, invasion, and proliferation [35]. Mutations in LCP1 have been reported as putative cancer drivers on the basis of whole-exome sequencing in independent benzo[a]pyrene (BaP)-derived post-stasis human mammary epithelial cell strains [36]. L-plastin is a protein that exerts a cell-protective effect against TNF cytotoxicity in breast cancer cell lines [37]. The actin-binding protein LCP1/L-PLASTIN has been verified to participate in CXCL12/CXCR4 signalling in breast cancer cells [38].

Therefore, we concluded that PDZK1, ATP2B4, and LCP1 might be key regulatory genes during the development of spontaneous IP in tree shrews. In addition, the DNA methylation of these genes may be a crucial functional regulator of tumorigenesis. Nevertheless, the epigenetic mechanisms participating in the modulation of these genes as well as genetic regions associated with the development of IP require further exploration.

Conclusions

Overall, our findings systematically demonstrated the changes in mRNA, lncRNA, and circRNA and allowed the characterization of the genome-wide DNA methylation profiles of IPs and normal mammary glands in tree shrews, providing valuable evidence for better understanding the development of mammary tumours. Our research also showed that DNA methylation influences the expression of genes associated with the development of spontaneous IP in tree shrews. Such analyses greatly improve the progress in exploring the characteristics of DNA methylation in the development of breast IP and provide new directions for the study of epigenetic markers and target genes for spontaneous mammary tumours.

Methods

Tissue specimens from tree shrews and their histology

In total, six tree shrews were obtained from the Institute of Medical Biology, Chinese Academy of Medical Science (IMB-CAMS), in Kunming. Animal experiments were approved by the animal ethics committee of IMB-CAMS (animal ethics approval number: DWSP201809003).

Both normal breast tissues and breast tumours were isolated after animals were euthanasia with intraperitoneal injection of pentobarbital sodium (100 mg/kg). The mammary tumours were surgically removed, and the size of each tumour was then measured, and its volume was calculated as $V = \pi/6 \times (a)^2(b)$, (where a and b represent the shortest and longest transverse diameters, respectively) [39]. A portion of the tissue samples was dissected, fixed in a 4% paraformaldehyde solution, embedded in paraffin, and stained with HE before being histologically diagnosed by pathologists. Another portion of the tissue samples was immediately frozen in liquid nitrogen and stored at -80 °C for subsequent experiments.

Library construction and WGBS

Genomic DNA was extracted from the samples via the cetyltrimethylammonium bromide (CTAB) method, and the DNA concentration and integrity were by using a NanoPhotometer® spectrophotometer (IMPLEN, CA, USA) and agarose gel electrophoresis, respectively. Then, DNA libraries for bisulphite sequencing were prepared. Briefly, genomic DNA was fragmented into 100–300 bp fragments by sonication (Covaris, Massachusetts, USA) and purified with a MiniElute PCR Purification Kit (QIAGEN, MD, USA). The fragmented DNAs were end repaired, and a single “A” nucleotide was added to the 3’ end of the blunt fragments. Then, the genomic fragments were ligated to methylated sequencing adapters. Fragments with adapters were bisulphite converted using the Methylation-Gold kit (ZYMO, CA, USA), and unmethylated cytosines were converted to uracils through sodium bisulphite treatment. Finally, the converted DNA fragments were PCR amplified and sequenced using an Illumina HiSeqTM 4000 PE 150 instrument.

Methylation level analysis

After data filtering, the acquired clean reads were mapped to the *Tupaia chinensis* (Chinese tree shrew) reference genome (TupChi_1.0) (GCF_000334495.1) using BSMAP software (v2.90) [40]. Then, a custom Perl script was applied to call methylated cytosines, and a correction algorithm was applied to the methylated cytosine results [41]. Methylation levels were calculated according to the methylated cytosine percentage in the global genome as well as in variant

regions of the genome for each sequence context (CG, CHG and CHH). To estimate variant methylation patterns in variant genomic regions, the methylation profiles of the flanking 2 kb regions as well as the gene body were plotted based on the average methylation levels for each window.

DMR analysis

To investigate DMRs between two groups, the minimum read coverage to call the methylation status of a base was set to 4. The DMRs in each sequence context (CG, CHG and CHH) were identified according to the following criteria: for CG, CHG, CHH and all C, the number in each window had to be ≥ 5 , 5, 15, and 20, respectively; the absolute value of the difference in the methylation ratio had to be ≥ 0.25 , 0.25, 0.15 and 0.2; and $q \leq 0.05$ was required for all. GO and KEGG pathway enrichment analyses were performed for DMR-related genes to explore the functional enrichment of genes influenced by DMRs.

Strand-specific library construction and whole-transcriptome sequencing

First, total RNA was extracted using TRIzol, and rRNAs were removed to retain mRNAs and ncRNAs. The enriched mRNAs and ncRNAs were fragmented into short fragments by using fragmentation buffer and reverse transcribed into cDNA with random primers. Second-strand cDNA was synthesized by using DNA polymerase I, RNase H, dNTP (dUTP instead of dTTP) and buffer. Then, the cDNA fragments were purified with a QiaQuick PCR extraction kit, end repaired, subjected to poly (A) addition, and ligated to Illumina sequencing adapters. Next, UNG (uracil-N-glycosylase) was used to digest the second-strand cDNA. Finally, the digested products were size selected by agarose gel electrophoresis, PCR amplified, and sequenced on the Illumina HiSeq TM 2000 platform.

Alignment with ribosomal RNA (rRNA) and the reference genome

After the filtering of clean reads, Bowtie2 (v2.2.8) [42] was applied for mapping reads to the rRNA database, removing the rRNA mapped reads. The remaining reads were used for subsequent assembly and analysis. Moreover, the rRNA-removed reads were mapped to TupChi_1.0 (GCF_000334495.1) by TopHat2 (v2.1.1) [43]. After alignment with TupChi_1.0, unmapped reads (or mapped very poorly) were aligned with Bowtie2 again; the reads that mapped to the genomes were removed, and the unmapped reads were collected for circRNA identification. In addition to identifying expressed genes and their quantitative expression, sequence alignment is helpful for discovering new transcripts.

Transcript reconstruction

The reconstruction of transcripts was performed with Cufflinks software [44], together with TopHat2, to identify new genes and new splice variants of known genes. The program reference annotation-based transcripts (RABT) was previously applied. Cufflinks constructed faux reads in terms of reference to make up for the effect of low-coverage sequencing. Finally, all reassembled fragments were aligned with reference genes, and similar fragments were discarded.

Novel transcript identification and annotation

To identify new transcripts, all reconstructed transcripts were aligned to TupChi_1.0. We applied the following parameters to identify predictable novel genes: length of transcript > 200 bp and exon number > 2. To acquire functional annotations of novel transcripts, alignment to the Nr, KEGG, and GO databases was performed.

CircRNA identification and database annotation

From both ends of the unmapped reads, 20-mers were extracted and aligned to TupChi_1.0 to identify unique anchor positions within splice sites. Moreover, anchor reads that aligned in the reversed orientation (head-to tail) and showed circRNA splicing were subjected to find_circ [45] analysis to identify circRNAs. Anchor alignments were extended such that the complete read aligned and the breakpoints were flanked by GU/AG splice sites. A candidate circRNA was called if it was supported by at least two unique back-spliced reads. Furthermore, the identified circRNAs were subjected to statistical analysis of their type and length distribution. Finally, circRNAs were subjected to BLAST searches against circBase [46] for annotation, and those that could not be annotated were defined as novel circRNAs.

LncRNA prediction and analysis

CNCI (v2) [47] and CPC [48] were applied to evaluate the protein-coding potential of novel transcripts according to default parameters. Novel transcripts were also mapped to the SwissProt database to obtain protein annotations. Those showing the intersection of neither protein-coding potential nor protein annotation results were chosen as lncRNAs. To investigate the interaction between antisense lncRNA and mRNA, RNAplex [49] was used to perform complementary correlation analysis. The program contains the ViennaRNA package [50] and predicts the best base pairing according to thermodynamic structure on the basis of the calculation of minimum free energy.

Quantification of transcript abundance and circRNA abundance

Transcript abundance was quantified by RSEM [51]. The fragments per kilobase of transcript per million mapped reads (FPKM) method was used to normalize transcript expression levels. In addition, the reads per million mapped reads (RPM) method was used to scale back-spliced junction reads to quantify circRNAs.

Analysis of differentially expressed transcripts and differentially expressed circRNAs

Differentially expressed mRNAs, lncRNAs and circRNAs were identified. To determine differentially expressed transcripts and circRNAs between two groups, the edgeR package was used. In each comparison, we identified mRNAs with a fold change ≥ 2 and FDR < 0.05 as DEGs and circRNAs/lncRNAs with a fold change ≥ 2 and P value < 0.05 as differentially expressed circRNAs/lncRNAs. The relevant coding RNAs were subjected to GO function and KEGG pathway enrichment analysis.

Correlation of DNA methylation and gene expression

To identify whether the DNA methylation level in DMRs affects gene expression between groups, genes were classified according to their genomic location, including the ± 2 kb flanking regions and gene body region. Spearman correlation analysis was used to statistically identify the relationships between gene expression and DNA methylation within the gene body and ± 2 kb flanking regions. $\rho < 0$ indicates a negative correlation and $\rho > 0$ a positive correlation. To investigate the underlying functions of DNA methylation responsible for differential gene expression, the common genes between the DMR-related genes and DEGs were analysed, and GO and KEGG pathway enrichment analyses were conducted for DEGs with DMRs.

GO and pathway enrichment analysis

GO enrichment analysis recognizes the key biological functions of genes by providing all GO terms that are significantly enriched in genes compared to the genomic background, in addition to filtering genes that correspond to biological functions. Moreover, genes generally interact with each other to carry out specific biological functions. Pathway-based analysis contributes to further identifying gene biological functions. KEGG is the primary public pathway-related database [52]. All genes were mapped to GO terms in the GO database, gene numbers were calculated for every term, and significantly enriched GO terms were identified. Pathway enrichment analysis revealed signal transduction or metabolic pathways that were significantly enriched in genes compared to the genome background. The calculated *P* values were subjected to FDR correction, taking a *P* value ≤ 0.05 as a threshold. GO terms/pathways meeting this criterion were defined as significantly enriched GO terms/pathways in genes.

GSEA

We carried out gene set enrichment analysis using GSEA software [53] to discern whether a set of genes enriched in distinct GO terms/pathways showed significant differences between two groups. In brief, we input a gene expression matrix and ranked genes by using the SinaltoNoise normalization program. Enrichment scores and *p* values were calculated using default parameters.

Abbreviations

IP: intraductal papilloma; WGBS: whole genome bisulphate sequencing; DMR: differentially methylated region; DMG: differentially methylated gene; DEG: differentially expressed gene; lncRNA: long noncoding RNA; circRNA: circular RNA; DCIS: ductal carcinoma in situ; RNA-seq: RNA sequencing; DNAm: DNA methylation; mC: methylated cytosine; TSS: transcription start site; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; BP: Biological Process; MF: Molecular Function; CC: Cellular Component; GSEA: gene set enrichment analysis; DElncRNAs: differentially expressed lncRNAs; DEcircRNAs: differentially expressed circRNAs;

GATA3: GATA-binding protein 3; hKLF: human Krüppel-like factor; PDZK1: PDZ Domain Containing 1; GWAS: Genome-wide association study; ERG: oestrogen regulated genes; ATP2B4: ATPase Plasma Membrane Ca²⁺ Transporting 4; PMCA4b: plasma membrane calcium ATPase isoform 4; LCP1: Lymphocyte cytosolic protein 1; CTAB: cetyltrimethylammonium bromide; rRNA: ribosomal RNA; RABT: reference annotation-based transcripts; FPKM: fragments per kilobase of transcript per million mapped reads; RPM: reads per million mapped reads; FDR: false discovery rate

Declarations

Acknowledgements

We thank Prof. Yunzhang Hu and Dr. Jiandong Shi (Institute of Medical Biology, Chinese Academy of Medical Science) for critical reading and constructive suggestions on this manuscript as well as Guangzhou Genedenovo Biotechnology Co., Ltd. for assisting in the sequencing and bioinformatics analyses.

Authors' contributions

D.J.J. and L.C.X. conceived and designed this study. K.D.X., L.C.X. L.N., W.W.G. and T.P.F. collected samples and conducted genome sequencing. and L.C.X. and H.Y.Y. performed the experiments and genome analysis and interpretation. L.C.X. wrote the paper. S.X.M. and D.J.J. contributed to the data interpretation and edited the paper.

Funding

This work was supported by Yunnan science and technology talent and platform program (No.2017HC019; 2018HB071); Yunnan Key Laboratory of Vaccine Research and Development on Severe Infectious Diseases (No.KF2015-01); the Yunnan health training project of high level talents (No.D-2018026); The Kunming Science and Technology innovation team (2019-1-R-24483). The funders had no role in the design of this study, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The WGBS and RNA-seq clean data obtained in this study were deposited in the NCBI Sequence Read Archive with accession numbers PRJNA663253 and PRJNA658225, respectively.

Declarations

Ethics approval and consent to participate

Animal care and all experimental protocols used in this study were approved by the Animal Ethics Committee of Animal Research in the Institute of Medical Biology, Chinese Academy of Medical Science prior to the beginning of this research (DWSP201809003). All of the animal procedures were approved by the ethical committee for Animal Research at the Institute of Medical Biology, Chinese Academy of Medical Science.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Karadeniz E, Arslan S, Akcay MN, Subasi ID, Demirci E: Papillary Lesions of Breast. *Chirurgia* 2016, 111(3):225-229.
2. Soran A, Balkan M, Harlak A, Evrensel T, Bonaventura M: Complicated solitary intraductal papilloma of the breast. *International journal of clinical practice* 2008, 62(1):160-161.
3. Jung SY, Kang HS, Kwon Y, Min SY, Kim EA, Ko KL, Lee S, Kim SW: Risk factors for malignancy in benign papillomas of the breast on core needle biopsy. *World journal of surgery* 2010, 34(2):261-265.
4. Eiada R, Chong J, Kulkarni S, Goldberg F, Muradali D: Papillary lesions of the breast: MRI, ultrasound, and mammographic appearances. *AJR American journal of roentgenology* 2012, 198(2):264-271.
5. Janecka JE, Miller W, Pringle TH, Wiens F, Zitzmann A, Helgen KM, Springer MS, Murphy WJ: Molecular and genomic data identify the closest living relative of primates. *Science* 2007, 318(5851):792-794.
6. Ye L, He M, Huang Y, Zhao G, Lei Y, Zhou Y, Chen X: Tree shrew as a new animal model for the study of lung cancer. *Oncology letters* 2016, 11(3):2091-2095.
7. Yang C, Ruan P, Ou C, Su J, Cao J, Luo C, Tang Y, Wang Q, Qin H, Sun W *et al*: Chronic hepatitis B virus infection and occurrence of hepatocellular carcinoma in tree shrews (*Tupaia belangeri chinensis*). *Virology journal* 2015, 12:26.
8. Tong Y, Hao J, Tu Q, Yu H, Yan L, Li Y, Lv L, Wang F, Iavarone A, Zhao X: A tree shrew glioblastoma model recapitulates features of human glioblastoma. *Oncotarget* 2017, 8(11):17897-17907.
9. Fan Y, Huang ZY, Cao CC, Chen CS, Chen YX, Fan DD, He J, Hou HL, Hu L, Hu XT *et al*: Genome of the Chinese tree shrew. *Nature communications* 2013, 4:1426.
10. Fan Y, Yu D, Yao YG: Tree shrew database (TreeshrewDB): a genomic knowledge base for the Chinese tree shrew. *Scientific reports* 2014, 4:7145.
11. Xia HJ, He BL, Wang CY, Zhang HL, Ge GZ, Zhang YX, Lv LB, Jiao JL, Chen C: PTEN/PIK3CA genes are frequently mutated in spontaneous and medroxyprogesterone acetate-accelerated 7,12-dimethylbenz(a)anthracene-induced mammary tumours of tree shrews. *European journal of cancer* 2014, 50(18):3230-3242.
12. Yang Y, Wu L, Shu XO, Cai Q, Shu X, Li B, Guo X, Ye F, Michailidou K, Bolla MK *et al*: Genetically Predicted Levels of DNA Methylation Biomarkers and Breast Cancer Risk: Data From 228 951 Women of European Descent. *Journal of the National Cancer Institute* 2020, 112(3):295-304.
13. Park JL, Jeon S, Seo EH, Bae DH, Jeong YM, Kim Y, Bae JS, Kim SK, Jung CK, Kim YS: Comprehensive DNA Methylation Profiling Identifies Novel Diagnostic Biomarkers for Thyroid Cancer. *Thyroid : official journal of the American Thyroid Association* 2020, 30(2):192-203.
14. Vrba L, Oshiro MM, Kim SS, Garland LL, Placencia C, Mahadevan D, Nelson MA, Futscher BW: DNA methylation biomarkers discovered in silico detect cancer in liquid biopsies from non-small cell lung cancer patients. *Epigenetics* 2020, 15(4):419-430.
15. Peng Y, Wu Q, Wang L, Wang H, Yin F: A DNA methylation signature to improve survival prediction of gastric cancer. *Clinical epigenetics* 2020, 12(1):15.
16. Xu Z, Sandler DP, Taylor JA: Blood DNA Methylation and Breast Cancer: A Prospective Case-Cohort Analysis in the Sister Study. *Journal of the National Cancer Institute* 2020, 112(1):87-94.
17. Lehmann U, Langer F, Feist H, Glockner S, Hasemeier B, Kreipe H: Quantitative assessment of promoter hypermethylation during breast cancer development. *The American journal of pathology* 2002, 160(2):605-612.
18. Bao Y, Wang L, Shi L, Yun F, Liu X, Chen Y, Chen C, Ren Y, Jia Y: Transcriptome profiling revealed multiple genes and ECM-receptor interaction pathways that may be associated with breast cancer. *Cellular & molecular biology letters* 2019, 24:38.
19. Mei Y, Yang JP, Lang YH, Peng LX, Yang MM, Liu Q, Meng DF, Zheng LS, Qiang YY, Xu L *et al*: Global expression profiling and pathway analysis of mouse mammary tumor reveals strain and stage specific dysregulated pathways in breast cancer progression. *Cell cycle* 2018, 17(8):963-973.
20. Kuang Y, Wang Y, Zhai W, Wang X, Zhang B, Xu M, Guo S, Ke M, Jia B, Liu H: Genome-Wide Analysis of Methylation-Driven Genes and Identification of an Eight-Gene Panel for Prognosis Prediction in Breast Cancer. *Frontiers in genetics* 2020, 11:301.
21. Liu X, Peng Y, Wang J: Integrative analysis of DNA methylation and gene expression profiles identified potential breast cancer-specific diagnostic markers. *Bioscience reports* 2020, 40(5).
22. Daino K, Nishimura M, Imaoka T, Takabatake M, Morioka T, Nishimura Y, Shimada Y, Kakinuma S: Epigenetic dysregulation of key developmental genes in radiation-induced rat mammary carcinomas. *International journal of cancer* 2018, 143(2):343-354.
23. Liu R, Shi P, Zhou Z, Zhang H, Li W, Zhang H, Chen C: Kruppel-like factor 5 is essential for mammary gland development and tumorigenesis. *The Journal of pathology* 2018, 246(4):497-507.
24. Shao M, Ge GZ, Liu WJ, Xiao J, Xia HJ, Fan Y, Zhao F, He BL, Chen C: Characterization and phylogenetic analysis of Kruppel-like transcription factor (KLF) gene family in tree shrews (*Tupaia belangeri chinensis*). *Oncotarget* 2017, 8(10):16325-16339.
25. Kocher O, Krieger M: Role of the adaptor protein PDZK1 in controlling the HDL receptor SR-BI. *Current opinion in lipidology* 2009, 20(3):236-241.

26. Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, Lush MJ, Maranian MJ, Bolla MK, Wang Q, Shah M *et al*: Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nature genetics* 2015, 47(4):373-380.
27. Ghosh MG, Thompson DA, Weigel RJ: PDZK1 and GREB1 are estrogen-regulated genes expressed in hormone-responsive breast cancer. *Cancer research* 2000, 60(22):6367-6375.
28. Haynes BP, Ginsburg O, Gao Q, Folked E, Afentakis M, Buus R, Quang LH, Thi Han P, Khoa PH, Dinh NV *et al*: Menstrual cycle associated changes in hormone-related gene expression in oestrogen receptor positive breast cancer. *NPJ breast cancer* 2019, 5:42.
29. Dunbier AK, Anderson H, Ghazoui Z, Folked EJ, A'Hern R, Crowder RJ, Hoog J, Smith IE, Osin P, Nerurkar A *et al*: Relationship between plasma estradiol levels and estrogen-responsive gene expression in estrogen receptor-positive breast cancer in postmenopausal women. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2010, 28(7):1161-1167.
30. Kim H, Abd Elmageed ZY, Ju J, Naura AS, Abdel-Mageed AB, Varughese S, Paul D, Alahari S, Catling A, Kim JG *et al*: PDZK1 is a novel factor in breast cancer that is indirectly regulated by estrogen through IGF-1R and promotes estrogen-mediated growth. *Molecular medicine* 2013, 19:253-262.
31. Varga K, Paszty K, Padanyi R, Hegedus L, Brouland JP, Papp B, Enyedi A: Histone deacetylase inhibitor- and PMA-induced upregulation of PMCA4b enhances Ca²⁺ clearance from MCF-7 breast cancer cells. *Cell calcium* 2014, 55(2):78-92.
32. Varga K, Hollosi A, Paszty K, Hegedus L, Szakacs G, Timar J, Papp B, Enyedi A, Padanyi R: Expression of calcium pumps is differentially regulated by histone deacetylase inhibitors and estrogen receptor alpha in breast cancer cells. *BMC cancer* 2018, 18(1):1029.
33. Curry MC, Luk NA, Kenny PA, Roberts-Thomson SJ, Monteith GR: Distinct regulation of cytoplasmic calcium signals and cell death pathways by different plasma membrane calcium ATPase isoforms in MDA-MB-231 breast cancer cells. *The Journal of biological chemistry* 2012, 287(34):28598-28608.
34. Baggott RR, Alfranca A, Lopez-Maderuelo D, Mohamed TM, Escolano A, Oller J, Ornes BC, Kurusamy S, Rowther FB, Brown JE *et al*: Plasma membrane calcium ATPase isoform 4 inhibits vascular endothelial growth factor-mediated angiogenesis through interaction with calcineurin. *Arteriosclerosis, thrombosis, and vascular biology* 2014, 34(10):2310-2320.
35. Pillar N, Polsky AL, Shomron N: Dual inhibition of ABCE1 and LCP1 by microRNA-96 results in an additive effect in breast cancer mouse model. *Oncotarget* 2019, 10(21):2086-2094.
36. Severson PL, Vrba L, Stampfer MR, Futscher BW: Exome-wide mutation profile in benzo[a]pyrene-derived post-stasis and immortal human mammary epithelial cells. *Mutation research Genetic toxicology and environmental mutagenesis* 2014, 775-776:48-54.
37. Janji B, Vallar L, Al Tanoury Z, Bernardin F, Vetter G, Schaffner-Reckinger E, Berchem G, Friederich E, Chouaib S: The actin filament cross-linker L-plastin confers resistance to TNF-alpha in MCF-7 breast cancer cells in a phosphorylation-dependent manner. *Journal of cellular and molecular medicine* 2010, 14(6A):1264-1275.
38. Inaguma S, Riku M, Ito H, Tsunoda T, Ikeda H, Kasai K: GLI1 orchestrates CXCR4/CXCR7 signaling to enhance migration and metastasis of breast cancer cells. *Oncotarget* 2015, 6(32):33648-33657.
39. Raa A, Stansberg C, Steen VM, Bjerkvig R, Reed RK, Stuhr LE: Hyperoxia retards growth and induces apoptosis and loss of glands and blood vessels in DMBA-induced rat mammary tumors. *BMC cancer* 2007, 7:23.
40. Xi Y, Li W: BSMAP: whole genome bisulfite sequence MAPping program. *BMC bioinformatics* 2009, 10:232.
41. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM *et al*: Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 2009, 462(7271):315-322.
42. Langmead B, Salzberg SL: Fast gapped-read alignment with Bowtie 2. *Nature methods* 2012, 9(4):357-359.
43. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL: TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* 2013, 14(4):R36.
44. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L: Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* 2012, 7(3):562-578.
45. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M *et al*: Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 2013, 495(7441):333-338.
46. Glazar P, Papavasileiou P, Rajewsky N: circBase: a database for circular RNAs. *Rna* 2014, 20(11):1666-1670.
47. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, Liu Y, Chen R, Zhao Y: Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic acids research* 2013, 41(17):e166.
48. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G: CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic acids research* 2007, 35(Web Server issue):W345-349.
49. Tafer H, Hofacker IL: RNAplex: a fast tool for RNA-RNA interaction search. *Bioinformatics* 2008, 24(22):2657-2663.
50. Lorenz R, Bernhart SH, Honer Zu Siederdisen C, Tafer H, Flamm C, Stadler PF, Hofacker IL: ViennaRNA Package 2.0. *Algorithms for molecular biology : AMB* 2011, 6:26.
51. Li B, Dewey CN: RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* 2011, 12:323.
52. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T *et al*: KEGG for linking genomes to life and the environment. *Nucleic acids research* 2008, 36(Database issue):D480-484.
53. 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.

Figures

Figure 1
Histological diagnosis of experimental samples. (A) Haematoxylin and eosin (H&E) staining results for breast intraductal papillomas in three tree shrews (magnification 400×). (B) H&E staining results for normal mammary tissue in three tree shrews (magnification 400×).

Figure 1
 Histological diagnosis of experimental samples. (A) Haematoxylin and eosin (H&E) staining results for breast intraductal papillomas in three tree shrews (magnification 400×). (B) H&E staining results for normal mammary tissue in three tree shrews (magnification 400×).

Figure 2

Genome-wide DNA methylation levels and trends. (A) The average ratio of DNA methylation types in the Control and IP groups. Blue, orange, and grey represent methylated (m) CG, mCHG, and mCHH, respectively. (B) CG methylation levels in the gene body and 2 kb upstream and downstream regions. Differently coloured lines represent different experimental repetitions.

Figure 2

Genome-wide DNA methylation levels and trends. (A) The average ratio of DNA methylation types in the Control and IP groups. Blue, orange, and grey represent methylated (m) CG, mCHG, and mCHH, respectively. (B) CG methylation levels in the gene body and 2 kb upstream and downstream regions. Differently coloured lines represent different experimental repetitions.

Figure 3

Differentially methylated genes (DMGs) associated with the development of IP. (A) Heatmap of DMGs located in differentially methylated regions (DMRs). (B) The number of DMGs in different regions of the genome in mammary gland tissues and IPs. (C) Annotation of DEGs with GO enrichment. Gene numbers and percentages are listed for each category. (D) KEGG pathway enrichment of DMGs. The rich factor is the ratio of the number of DEGs associated with the pathway term to the number of all genes associated with the pathway term. The Q value is corrected to range from 0 to 1. Only the top 20 enriched pathway terms are displayed here.

Figure 3

Differentially methylated genes (DMGs) associated with the development of IP. (A) Heatmap of DMGs located in differentially methylated regions (DMRs). (B) The number of DMGs in different regions of the genome in mammary gland tissues and IPs. (C) Annotation of DEGs with GO enrichment. Gene numbers and percentages are listed for each category. (D) KEGG pathway enrichment of DMGs. The rich factor is the ratio of the number of DEGs associated with the pathway term to the number of all genes associated with the pathway term. The Q value is corrected to range from 0 to 1. Only the top 20 enriched pathway terms are displayed here.

Figure 4

Differentially expressed mRNAs and gene set enrichment analysis (GSEA). (A) Volcano plot of differentially expressed mRNAs. (Red points represent upregulated mRNAs; green points represent downregulated mRNAs; black points represent unchanged mRNAs). (B) GSEA of up- or downregulated gene sets in IP compared to those in normal mammary tissues. Enrichment plots for the target gene sets (FDR < 0.05). NES, normalized enrichment score; FDR, false discovery rate.

Figure 4

Differentially expressed mRNAs and gene set enrichment analysis (GSEA). (A) Volcano plot of differentially expressed mRNAs. (Red points represent upregulated mRNAs; green points represent downregulated mRNAs; black points represent unchanged mRNAs). (B) GSEA of up- or downregulated gene sets in IP compared to those in normal mammary tissues. Enrichment plots for the target gene sets (FDR < 0.05). NES, normalized enrichment score; FDR, false discovery rate.

Figure 5

Differentially expressed lncRNAs associated with the development of IP. (A) Volcano plot of differentially expressed lncRNAs. (Red points represent upregulated lncRNAs; green points represent downregulated lncRNAs; black points represent unchanged lncRNAs.) Gene Ontology (GO) and pathway analyses of target genes of significant differentially expressed lncRNAs. Bar plots (B, D, F.) display enriched GO terms; bubble charts (C, E, G.) represent KEGG pathways. The plots show significantly enriched GO terms and pathways ($P < 0.05$).

Figure 5

Differentially expressed lncRNAs associated with the development of IP. (A) Volcano plot of differentially expressed lncRNAs. (Red points represent upregulated lncRNAs; green points represent downregulated lncRNAs; black points represent unchanged lncRNAs.) Gene Ontology (GO) and pathway analyses of target genes of significant differentially expressed lncRNAs. Bar plots (B, D, F.) display enriched GO terms; bubble charts (C, E, G.) represent KEGG pathways. The plots show significantly enriched GO terms and pathways ($P < 0.05$).

Figure 6

Differentially expressed circRNAs associated with the development of IP. (A) Heat map of differentially expressed circRNAs. (Red columns represent upregulated genes in IP tissues, while blue columns represent downregulated genes.) (B) Parental genes of differentially expressed circRNAs are classified into cellular components, molecular function and biological processes according to GO term. (C) KEGG pathway enrichment of parental genes of differentially expressed circRNAs. The size of the circle represents the gene number, and the colour represents the P value.

Figure 6

Differentially expressed circRNAs associated with the development of IP. (A) Heat map of differentially expressed circRNAs. (Red columns represent upregulated genes in IP tissues, while blue columns represent downregulated genes.) (B) Parental genes of differentially expressed circRNAs are classified into cellular components, molecular function and biological processes according to GO term. (C) KEGG pathway enrichment of parental genes of differentially expressed circRNAs. The size of the circle represents the gene number, and the colour represents the P value.

Figure 7

Integrative analysis between DNA methylation and transcriptome data. (A) Chart of the correlation analysis of all gene expression levels and DNA methylation levels in the groups. (B) Correlation analysis of the gene methylation level and gene expression level in the Control and IP groups. (C) Heat map analysis of gene methylation and expression levels in IPs compared to normal mammary glands.

Figure 7

Integrative analysis between DNA methylation and transcriptome data. (A) Chart of the correlation analysis of all gene expression levels and DNA methylation levels in the groups. (B) Correlation analysis of the gene methylation level and gene expression level in the Control and IP groups. (C) Heat map analysis of gene methylation and expression levels in IPs compared to normal mammary glands.

Supplementary Files

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

- [Additionalfile1.zip](#)
- [Additionalfile1.zip](#)
- [AuthorChecklistFull.pdf](#)
- [AuthorChecklistFull.pdf](#)