British Journal of Cancer

ARTICLE OPEN

Check for updates

Genetics and Genomics

An integrative approach for exploring the nature of fibroepithelial neoplasms

Jihui Yun[1,2,13], Woohang Heo[3,13], Eun-Shin Lee[4,13], Deukchae Na[5], Wonyoung Kang[6,7], Jinjoo Kang[7], Jeesoo Chae[1,2], Dakyung Lee[1,2], Woochan Lee[1,2], Jinha Hwang[1,2], Tae-Kyung Yoo[4], Bok Sil Hong[8], Hye-Youn Son[8], Dong-Young Noh[4,9,10], Charles Lee[6,7,11], Hyeong-Gon Moon[1,4,9,10,14]✉ and Jong-Il Kim[1,2,9,12,14]✉

© The Author(s) 2022

BACKGROUND: Malignant phyllodes tumour (MPT) is a rare breast malignancy with epithelial and mesenchymal features. Currently, there are no appropriate research models or effective targeted therapeutic approaches for MPT.
METHODS: We collected fresh frozen tissues from nine patients with MPT and performed whole-exome and RNA sequencing. Additionally, we established patient-derived xenograft (PDX) models from patients with MPT and tested the efficacy of targeting dysregulated pathways in MPT using the PDX model from one MPT.
RESULTS: MPT has unique molecular characteristics when compared to breast cancers of epithelial origin and can be classified into two groups. The PDX model derived from one patient with MPT showed that the mouse epithelial component increased during tumour growth. Moreover, targeted inhibition of platelet-derived growth factor receptor (PDGFR) and phosphoinositide 3-kinase (PI3K)/mammalian target of rapamycin (mTOR) by imatinib mesylate and PKI-587 showed in vivo tumour suppression effects.
CONCLUSIONS: This study revealed the molecular profiles of MPT that can lead to molecular classification and potential targeted therapy, and suggested that the MPT PDX model can be a useful tool for studying the pathogenesis of fibroepithelial neoplasms and for preclinical drug screening to find new therapeutic strategies for MPT.

British Journal of Cancer (2023) 128:626–637; https://doi.org/10.1038/s41416-022-02064-2

BACKGROUND

Breast fibroepithelial tumours, including fibroadenomas and phyllodes tumours, are biphasic neoplasms characterised by proliferation of epithelial and stromal cells of the mammary glands [1]. Phyllodes tumours comprise <1% of all primary breast neoplasms, and up to one-third of phyllodes tumours are classified as malignant phyllodes tumours (MPTs) based on histologic findings, such as degree of mitosis and extent of stromal growth [1]. Up to 67% of patients with MPT develop local recurrences, and 21% experience distant metastasis [2].

The current standard treatment for MPT is surgical excision of the tumour with sufficient resection margins. The addition of postsurgical radiation therapy did not result in substantial improvement in overall survival, and the benefit of adjuvant systemic therapy has not been tested in randomised clinical trials [3, 4]. Owing to the rare nature of the neoplasm and our incomplete understanding of the molecular characteristics of the disease, there are no effective systemic or targeted treatment options for MPT despite high rates of recurrence.

Recent studies have improved our understanding of the molecular characteristics of MPT [4, 5]. Scientists have identified recurrent chromosomal alterations [6, 7], gene expression characteristics [7, 8], and somatic mutations [9, 10]. However, many have primarily focused on genomic profiling of a wide range of phyllodes neoplasms, and thus included large proportions of benign and borderline phyllodes tumours. Furthermore, few studies have addressed the potential in vivo application of targeted therapies in MPT, owing to the lack of available cell lines or animal models.

Here, we presented the genomic and transcriptomic characteristics of MPT using various tumour tissues and suggested gene expression-based molecular subtypes of MPT. Additionally, we elucidated the driver and microenvironment relationship in this biphasic neoplasm using a patient-derived xenograft (PDX) model

[1]Genomic Medicine Institute, Medical Research Center, Seoul National University, Seoul, Republic of Korea. [2]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Republic of Korea. [3]Interdisciplinary Program on Tumor Biology, Seoul National University College of Medicine, Seoul, Republic of Korea. [4]Department of Surgery, Seoul National University Hospital, Seoul, Republic of Korea. [5]Ewha Institute of Convergence Medicine, Ewha Womans University Mokdong Hospital, Seoul, Republic of Korea. [6]The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut, USA. [7]Department of Life Sciences, Ewha Womans University, Seoul, Republic of Korea. [8]Center for Medical Innovation, Seoul National University Hospital, Seoul, Republic of Korea. [9]Cancer Research Institute, Seoul National University, Seoul, Republic of Korea. [10]Department of Surgery, Seoul National University College of Medicine, Seoul, Republic of Korea. [11]Precision Medicine Center, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, People's Republic of China. [12]Department of Biochemistry and Molecular Biology, Seoul National University College of Medicine, Seoul, Republic of Korea. [13]These authors contributed equally: Jihui Yun, Woohang Heo, Eun-Shin Lee. [14]These authors jointly supervised this work: Hyeong-Gon Moon, Jong-Il Kim. ✉email: moonhg74@snu.ac.kr; jongil@snu.ac.kr

Received: 27 January 2022 Revised: 9 November 2022 Accepted: 10 November 2022
Published online: 16 December 2022

derived from MPT. Finally, we proposed novel therapeutic approaches for MPT by targeting the platelet-derived growth factor receptor (PDGFR) and phosphoinositide 3-kinase (PI3K)/mammalian target of rapamycin (mTOR) pathways.

METHODS

Patients and biospecimens

Frozen tumour tissues of MPTs were obtained from the Institutional Review Board (IRB)-approved breast cancer biospecimen repository at Seoul National University Hospital (IRB no. 1509-032-702). Tissues were obtained from surgical specimens during curative surgeries. Matched germline DNA was obtained from the peripheral blood of the patient. Formalin-fixed, paraffin-embedded (FFPE) MPT tissues were obtained from the Department of Pathology, Seoul National University Hospital (IRB no. 1402-054-555).

Patient-derived xenograft models

To develop the PDX models, surgically resected tumour tissues were minced into approximately 2 mm pieces and transplanted into the fourth mammary fat pad of 8-week-old female NOD/SCID/IL-2γ-receptor null (NSG, Jackson Lab, USA) mice. When the MPT PDX tumour exceeded 500 mm^3 after transplantation into the mouse mammary fat pad, it was excised and cut into 2–3 mm^3 pieces. Subsequently, a hole was made in the mammary fat pad of the mouse to be transplanted, and a piece of PDX tumour was placed in the hole. Thus, MPT PDX models were established in several passages (P1 and P2). Tumour volumes and body weight of mice were measured once or twice a week.

Whole-exome sequencing (WES) and data processing

Exome libraries captured using the SureSelect Human All Exon V4 or V5 kit (Agilent, Santa Clara, CA, USA) for WES were sequenced on an Illumina platform. Most bioinformatic analyses of sequencing data were performed using the computing server at the Genomic Medicine Institute Research Service Center. We aligned the DNA sequence reads to the human reference genome (GRCh37) or a combined reference genome of human (GRCh37) and mouse (mm10) using Burrows–Wheeler Aligner (BWA) [11]. Sorting of reads and marking of PCR-duplicated reads were performed using Picard (http://broadinstitute.github.io/picard/). For samples using a combined reference genome of human and mouse, we removed mouse sequence reads according to additional processing procedures [12]. Thereafter, we performed preprocessing procedures for bam files, including local realignment around insertions/deletions (indels) and base recalibration, according to the Genome Analysis Toolkit (GATK) best practices document [13, 14].

Mutational landscape analysis using WES data

We discovered somatic mutations using MuTect [15] for single-nucleotide variants (SNVs) and GATK IndelGenotyper V2 [14] for indels. ANNOVAR [16] was used to annotate the somatic mutations. We used the following criteria to identify true positive variants in the exome data: alt allele depth ≥4 and "MuTect=PASS". For indels, total depth ≥8, alt allele depth ≥6, and alt allele frequency ≥0.1. Only coding or splice-site variants were maintained.

The number of variants that passed the above criteria was described as the number of "Total" variants in Supplementary Table S1. Somatic mutation burden was calculated by the number of "Total" variants per megabase for exome. We performed an analysis to identify mutational processes using deconstructSigs [17] based on mutational signatures in COSMIC [18] with default settings or with additional normalisation. SigMA can be used to detect mutational signatures in sequencing data with a small number of mutations [19]. We used this tool to more accurately detect signature 3, which is frequently found in breast cancer, using a web-based interface (http://compbio.med.harvard.edu/sigma/). As it is difficult to filter germline mutations in one sample without a matched normal sample, the "mpt-06" sample was excluded from the somatic mutation burden and mutational signature analysis.

Additional filters were applied to identify more noteworthy variants. Following removal of variants annotated as "synonymous" or "unknown" by ANNOVAR [16], we only retained variants that rarely exhibited or did not report in the general population database, including the 1000 Genomes Project, Exome Aggregation Consortium, and Exome Sequencing Project. The number of variants that passed the additional filter was described as the number of "Noteworthy" variants in Supplementary Table S1. To identify somatic mutations in one tumour sample without a matched blood sample (mpt-06), the variants considered germline variants by the Korean germline variant database [20] were removed from the variants identified using the MuTect tumour-only mode.

MutationMapper [21, 22] was used for lollipop plot visualisation of specific mutations. Additional interpretation of mutations was performed by MutationMapper and Oncotator [23]. Cancer-related genes were defined on the basis of the COSMIC Cancer Gene Census [24]. We surveyed clinically relevant variants to discover potential therapeutic targets based on the TARGET database v3 from the Broad Institute (https://software.broadinstitute.org/cancer/cga/target).

Somatic copy number alterations (SCNAs) analysis

We discovered SCNAs from exome sequencing data using EXCAVATOR [25], the R package DNAcopy, and CNVkit [26]. After generation of read coverage for all samples using CoNIFER [27], we calculated the log2 ratio of tumour and matched blood samples, and used these values as input for DNAcopy analysis. For one sample without a matched blood sample, we utilised a pool of blood samples that used the same exome capture kit as the baseline for DNAcopy and CNVkit analysis. The SCNAs identified from the DNAcopy are shown in Supplementary Fig. S2a. To identify important SCNAs, we detected genes in regions with copy numbers >4 or genes where mutations appeared among cancer-related genes, or recurrently variable regions from results by EXCAVATOR.

RNA sequencing (RNA-seq) and data processing

Library construction for RNA-seq was performed using the TruSeq RNA Sample Prep Kit v2 for most samples, TruSeq RNA Access Library Prep kit for FFPE, or TruSeq Stranded mRNA Prep kit (Illumina, San Diego, CA, USA) for several MPT PDX samples. Sequencing of cDNA libraries was performed using the Illumina platform. We used a human genome reference and a combined reference for human and mouse to align sequence reads by spliced transcripts alignment to a reference (STAR) aligner [28]. The following procedures were conducted referring to the best practices workflow for RNA-seq of GATK. The number of reads for each Ensembl annotated gene was quantified from pre-processed data using HTSeq-count [29]. We applied parameters that were considered appropriate for each sequencing data. Raw read counts from HTSeq-count were converted to fragments per kilobase of exon per million reads mapped (FPKM) values using the rpkm() function of the R package edgeR [30].

To find *MED12* mutations from RNA-seq data of FFPE, we manually examined only known mutation sites in *MED12* exon 2, previously found in phyllodes tumour [10] from the bam files using the Integrative Genomics Viewer (IGV).

Differential gene expression analysis

Differentially expressed genes (DEGs) were identified using the R package DESeq2 [31]. We only retained genes with an adjusted p value < 0.05, log2 fold-change ≥1, and average of expression (FPKM) across all samples ≥1 to identify more important DEGs from the results of differential gene expression analysis. To identify noteworthy enriched pathways from DEGs, we used the DAVID Bioinformatics Database functional annotation tool (http://david.abcc.ncifcrf.gov/). Statistical significance was set at $p < 0.05$. The ClueGO [32] Cytoscape plugin was used to visualise the enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways [33]. Significance in ClueGO was determined by the p-value corrected with Bonferroni step-down.

Visualisation of transcriptome profiles

After adding 1 to the FPKM of each gene, these values were log transformed and mean-centred. The adjusted expression values were hierarchically clustered based on uncentered correlation and average linkage by cluster 3.0 [34]. Visualisation of the heatmap was performed using Java Treeview [35].

Read counts of each gene from HTSeq-count were transformed by the variance stabilising transformation (vst) function using the R package DESeq2 [31]. After determining the variance for each gene, the 1000 most variable genes were used for principal component analysis (PCA). We generated a PCA plot based on the first two principal components (PC1 and PC2) using R package's ggplot2 or Prism (GraphPad).

Molecular classification of MPT samples

To identify classifier genes which divide MPT samples into the "Epithelial" and "Fibrous" subtypes, we applied the criteria below and selected 32 genes among noteworthy DEGs identified in differential gene expression analysis between the two subtypes of nine MPT fresh frozen (FF) tissues: (1) protein-coding gene, (2) log2 of fold-change among genes overexpressed in "Epithelial" ≥8, log2 of fold-change among genes overexpressed in "Fibrous" ≥2, (3) minimum difference of expression (FPKM) between the two groups of MPT FF tissues ≥1, and (4) mean expression (FPKM) in 28 MPT FFPE tissues ≥2.

Histological analysis

MPT PDX tumours were resected once the tumour volume reached 200, 500, and 1000 mm^3, and paraffin-embedded blocks were prepared and mounted on microscope slides. To quantify the ratio of epithelial cells in these MPT PDX tumours, an image analysis program "QuPath" was used. To discriminate between human mesenchymal cells and mouse epithelial cells histologically, PDX tissue slides were rehydrated, and immunostaining was performed with anti-human HLA class 1 (Abcam, 1:800). For human and mouse pan-centromeric FISH followed by anti-FITC staining, we used Star*FISH human and mouse pan-centromeric probes-FITC (Cambio) according to the manufacturer's protocol. Briefly, after rehydration, slides were incubated with pepsin solution and quenched with glycine solution. Following washing with PBS, slides were post-fixed in paraformaldehyde solution. Thereafter, tissue slides were dehydrated, stained with chromosome paint overnight, and subsequently washed with formamide, stringency, detergent wash solution, and PBS sequentially, followed by HRP-conjugated anti-FITC (Abcam, 1:50) immunostaining for 1 h at room temperature. After mounting, slides were observed under a light microscope.

Single-cell RNA sequencing (scRNA-seq) data processing and analysis

Raw 3′ scRNA-seq data were aligned using Cell Ranger [36] to generate a count matrix. Sequencing reads were aligned to the human mouse (GRCh38-mm10) chimeric reference genome. For the analysis of tumour cells treated with imatinib, PKI-587, and vehicle, we removed doublets in each group using DoubletFinder [37]. SoupX [38] was used to remove technical ambient RNA. Expression matrices were further processed in R using Seurat [39]. The "Subset" function was applied to separate mouse and human cells. After removing unwanted cells, the "NormalizeData" function with the "LogNormalize" method was used to normalise filtered gene-barcode matrices, and the top 2,000 features were found using the "FindVariableFeatures" function with the "vst" method. Scaling was performed on the gene expression matrices using the "ScaleData" function. Thereafter, we performed PCA on scaled data and used the first 30 principal components (PCs) for further analysis. Harmony was used to integrate the data. The top 30 PCs were used as input to the UMAP dimension reduction. Nearest-neighbour graphs were calculated using "FindNeighbors" with the top 30 PCs and "FindClusters" function applied with a resolution of mouse = 0.8 and human = 0.8, respectively. Harmony [40] was run on the PCA matrix using the default parameters with a library batch. Human cells containing at least 1% mouse reads were removed, and filtered human cells were considered the tumour cell population. The Seurat "AddModuleScore" was used to calculate the PI3K pathway score. Gene sets with *PIK3CA*, *PIK3CB*, *AKT1*, *MTOR*, and *RPS6KB1* were used to calculate module scores.

Human and mouse cells separation in MPT PDX

To separate human and mouse cells from MPT PDX tumours, we used the MACS mouse cell depletion kit (Miltenyi Biotec) designed for isolating human cell populations in xenograft tumours using magnetically labelled beads targeting mouse cells. To dissociate the MPT PDX tumour, we resected the PDX tumour at 1000 mm^3 and dissociated it using a tumour dissociation kit with a gentle MACS dissociator (Miltenyi Biotec) according to the manufacturer's protocol. Dissociated cells were labelled with mouse-specific magnetic beads and separated using LS columns to obtain unlabelled flowing human cells. Following human cell isolation, bead-labelled mouse cells were detached from the LS column.

To verify whether cells were successfully separated by the host species, RNA was extracted from cells collected from each human and mouse using TRIzol reagent. Using a cDNA synthesis kit (Takara), cDNA was synthesised, and qRT-PCR was performed using the human GAPDH primer (F:5′-GAG TCC ACT GGC GTC TTC-3′ R:5′-GGA GGC ATT GCT GAT GAT C-3′) and mouse GAPDH primer (F:5′-GCC TTC CGT GTT CCT ACC-3′ R:5′-GCC TGC TTC ACC ACC TTC-3′).

Drug efficacy test in the MPT PDX

Drug treatment was initiated after the tumours reached approximately 200 mm^3. Mice were randomly divided into three treatment groups consisting of five mice in each group: (1) vehicle (0.5% methylcellulose, daily per oral), (2) imatinib mesylate (100 mg/kg, 0.5% methylcellulose, daily per oral), and (3) PKI-587 (25 mg/kg, dissolved in 5% dextrose, 0.3% lactic acid, weekly intravenous injection) for 3 weeks. The tumour volume and body weight of the mice were measured once or twice a week. Volume was calculated as (length × $width^2$)/2.

Western blotting

For protein extraction, a piece of tumour tissue (approximately 1 mm^3) was homogenised in ice-cold M-PER containing a protease and phosphatase inhibitor cocktail (Thermo Scientific) using a disposable homogeniser (BioMasher II tissue homogeniser, Nippi), briefly sonicated, and centrifuged at 14,000 × *g* for 20 min at 4 °C. Total proteins were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and transferred onto a polyvinylidene difluoride membrane. Following blocking, membranes were probed with specific antibodies against phospho-Akt (Ser473), Akt, phospho-mTOR (Ser2448), mTOR, phospho-MAPK (Thr202/Tyr204), MAPK, phospho-PDGFRa (Tyr849)/PDGFRb (Tyr857), PDGFRb (Cell Signaling Technology), and β-actin (Santa Cruz Biotechnology), followed by treatment with horseradish peroxidase-conjugated secondary antibodies. Target proteins were detected using Amersham Imager 600 (GE Healthcare) developed with an enhanced chemiluminescent substrate (Thermo Scientific).

Utilisation of in-house data

We utilised WES data on 18 invasive ductal carcinoma (IDC) samples and transcriptome data from 32 non-tumour breast tissues, 41 IDC tissues, 46 PDX tissues derived from IDC, and 6 normal mouse fat pads. The sequencing data of samples were generated similar to the samples used in this study for a separate project, to characterise genomic profiles of primary breast tumour and PDX tumour tissues, which will be presented separately (IRB no. 1402-054-555).

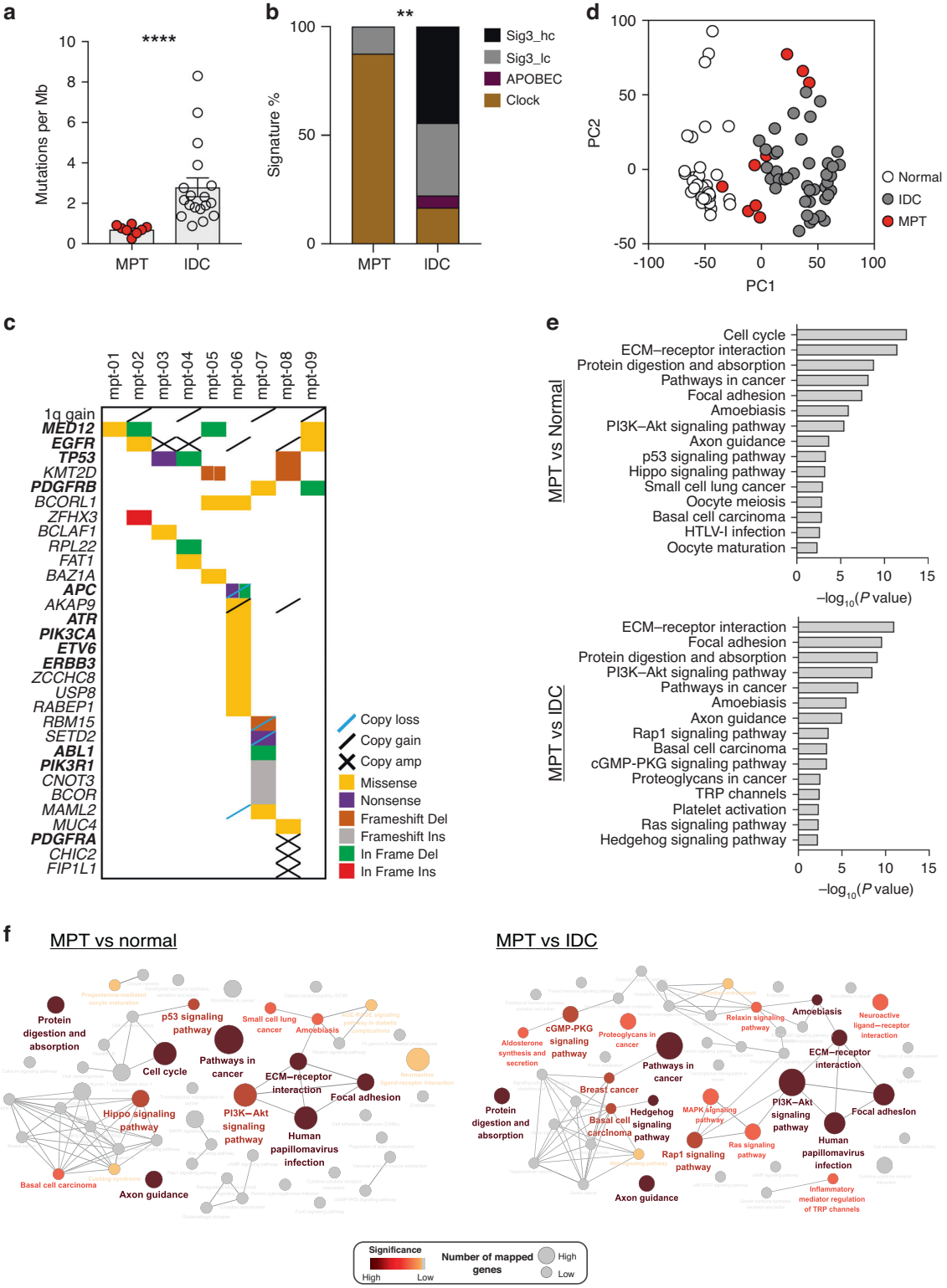
Statistical analyses

Most statistical calculations were performed using Prism version 8.0 for Windows (GraphPad software, La Jolla, CA, USA, www.graphpad.com). To identify statistically significant differences, the Mann–Whitney test was used for most statistical analyses. Fisher's exact test was used for comparison of the frequency of signature 3 between MPT and IDC and the frequency of *MED12* mutation according to molecular subtype. The chi-square test was used for comparison of degree of stromal overgrowth according to the molecular subtype and comparison of cellular composition upon drug treatment. The Kruskal–Wallis test was used for comparison of expression level of cell-type markers based on mouse transcriptome from MPT, IDC PDX, and normal fat pads. The Wilcoxon test was used for tumour growth comparison of drug-treated "MX-99" xenograft model. The Student's *t* test was used for comparison of PI3K activity upon drug treatment. Survival curves were created using the Kaplan–Meier method and compared using the log-rank test. All data are presented as the mean ± standard deviation (SD). *p* values < 0.05 were considered significant and indicated by $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$.

RESULTS

Genomic and transcriptomic characteristics of MPT of the breast

We obtained fresh frozen (FF) tissues from nine MPT patients in our biorepository and performed exome and transcriptome sequencing (Supplementary Table S1). Compared to exome sequencing data generated from an independent cohort of 18 IDC tissues, the MPT showed a significantly lower incidence of somatic mutations (0.7/Mb for MPT and 2.79/Mb for IDC) (Fig. 1a and Supplementary Fig. S1a). The most prevalent somatic



mutation signature, as proposed by Alexandrov et al. [41], of MPT was the C · G → T · A substitutions resulting from endogenous DNA damage (Supplementary Fig. S1b–d). The low incidence of the *BRCA*-related mutational signature (signature 3) and APOBEC editing-related signatures (signatures 2 and 13) suggested that the mutational process of MPT is distinct from that of IDC of the breast [42] (Fig. 1b and Supplementary Fig. S1d). The incidence of structural variations and somatic mutations in cancer-related genes in nine MPT is shown in Fig. 1c (Supplementary Data 1). Consistent with previous reports [4, 9, 43–45], we observed a high rate of *MED12* mutations in our series (4/9, 44.4%). Additionally, somatic mutations in *PIK3CA*, *PIK3R1*, *PDGFRB*, *SETD2*, and *TP53* were observed among the nine MPT. Regarding structural variations, we observed heterogeneous structural alterations, including 1q gain in four cases (44.4%) and EGFR amplifications in four cases (44.4%) (Fig. 1c and Supplementary Fig. S2a).

Fig. 1 Unique molecular characteristics of malignant phyllodes tumours (MPTs). a Prevalence of somatic mutations, measured by the number of somatic mutations per Mb of covered target sequence, in MPT compared to that in invasive ductal carcinoma (IDC) (MPT, $n = 8$; IDC, $n = 18$; Mann–Whitney test). **b** Comparison of the prevalence of signatures between MPT and IDC identified by SigMA (MPT, $n = 8$; IDC, $n = 18$; Fisher's exact test). "Sig3 hc" means signature 3 with high confidence, and "Sig3 lc" means signature 3 with low confidence. **c** Somatic mutations and copy number alterations in cancer-related genes and specific regions in the nine MPT. Genes associated with clinical action are denoted in bold font, based on TARGET v3 from the Broad Institute. Genes marked with "copy amp" indicate genes with a copy number ≥5. **d** PCA plot showing unique gene expression profiles of MPT compared to normal breast and IDC tissues. **e** KEGG pathways enriched with upregulated genes in MPT compared to normal breast tissues (top) or IDC tissues (bottom). **f** Cytoscape network analysis showing significantly altered pathways in MPT compared to normal breast tissues (left) or IDC tissues (right). The colour-filing nodes reflect statistical significance of the term (pathway). Node size refers to the number of genes associated with the term among the differentially expressed genes. Terms with high significance were clearly labelled. $^{**}p < 0.01$, $^{****}p < 0.0001$.

The transcriptome sequencing data of the nine MPT were compared with 32 normal breast tissues and 41 IDC tissues to determine the gene expression characteristics of the MPT. PCA using gene expression profiles revealed that MPT have unique characteristics compared to IDC or normal breast tissues (Fig. 1d). Pathway enrichment analysis using DEGs showed that, in addition to the cell cycle and pathways in cancer, genes involved in extracellular matrix (ECM) interactions and the PI3K signalling pathway were upregulated in MPT compared to normal breast tissues. The genes involved in ECM interactions and PI3K signalling were also upregulated in MPT compared to IDC (Fig. 1e). Alternatively, when compared to normal breast tissues, gene sets related to metabolism and the PPAR signalling pathway were downregulated in MPT, and when compared to IDC, genes involved in cell adhesion molecules and tight junctions were downregulated in MPT (Supplementary Fig. S2b). Interaction network analysis using Cytoscape [46] further revealed that ECM-related pathways and PI3K-Akt signalling pathways were significantly enriched in MPT (Fig. 1f). These findings suggested that MPT has unique genomic and transcriptomic characteristics that differ from those of the IDC of the breast.

Epithelial and mesenchymal features determine the clinically-relevant molecular subtypes of MPT

Unsupervised clustering using gene expression data suggested that MPT can be classified into two distinct molecular subtypes (Fig. 2a and Supplementary Fig. S3a). We identified DEGs between the two molecular subtypes of MPT (Supplementary Data 2 and Supplementary Fig. S3b). One subtype was enriched with genes involved in protein digestion and absorption, ECM-receptor interaction, focal adhesions, and axon guidance (Fig. 2b). Many mesenchyme-related genes, such as various types of collagens [47], were significantly upregulated (Fig. 2b). The other subtype showed upregulation of genes involved in tight junctions, cell metabolism, cell adhesion molecules, and various immune-related processes (Fig. 2c). This subtype showed upregulation of epithelial markers such as *CDH1*, *CLDN3*, *CLDN4*, *CLDN7*, and *OCLN* (Fig. 2c) [48–50]. In addition to examining expression levels of individual genes, we determined global expression patterns of genes associated with cell adhesion molecules (KEGG ID: hsa04514), ECM–receptor interaction (KEGG ID: hsa04512) from the KEGG database (http://www.genome.jp/kegg/) [33], and the epithelial-to-mesenchymal transition (EMT) [51], as the above gene sets could represent the epithelial or stromal spectrum of the MPT. Unsupervised clustering based on expression profiles of genes in the three gene sets resulted in the same classification of the nine tumours (Fig. 2d). These findings indicated that human MPT can be classified into two distinct molecular subtypes based on their mesenchymal and epithelial gene expression characteristics. We named the observed subtypes as "Epithelial" and "Fibrous", respectively which reflect the biphasic nature of fibroepithelial neoplasms.

We identified that mutations in *MED12* were found only in the "Epithelial" subtype (4/6 for "Epithelial," 0/3 for "Fibrous"; Fisher's exact test, $p = 0.1667$; Fig. 1c). Notably, MPT samples belonging to the "Fibrous" subtype showed higher expression levels of *PDGFRB* compared to those of MPT samples belonging to the "Epithelial" subtype, although the difference was not significant (Supplementary Fig. S3c). As it is known that the expression level of *PDGFRB* increases with increasing grade and that the frequency of *MED12* mutation decreases with increasing grade in the spectrum from benign to malignant phyllodes tumour [44, 52], we speculated that "Fibrous" would be a more advanced group compared to "Epithelial" within the malignant histologic category of phyllodes tumours. We performed RNA-seq using additional 28 MPT FFPE tissues to investigate the impact of gene expression features of molecular subtypes of MPT on clinical parameters of patients with MPT. We developed a gene expression signature consisting of 32 genes that could be used to classify MPT samples into two distinct groups and applied it to classify MPT FFPE samples (Supplementary Data 2 and Fig. 3a). We confirmed that the MPT FFPE samples were classified into two clusters based on the 32-gene expression signature (Fig. 3a). MPT FFPE samples also tended to form clusters in the PCA plot according to the group classified by the 32-gene expression signature (Supplementary Fig. S3d). To determine the degree of epithelial gene expression in tumours in each subgroup, we used xCell [53], a gene signature-based analytical tool that has been developed to identify specific cellular types from bulk RNA-seq data sets. As shown in Fig. 3b, xCell-based epithelial gene expression was significantly associated with molecular subtypes in MPT FF and FFPE tissues. The subgroup enriched with mesenchymal-related pathways showed significantly lower levels of xCell-based epithelial genes (Fig. 3b). We confirmed that mutations in *MED12* exon 2, which are frequently mutated in phyllodes tumours, including mutations in codon 44 (p.G44) of *MED12*, appear only in the "Epithelial" group in FFPE tissues as in FF tissues (3/15 for "Epithelial," 0/13 for "Fibrous"; Fisher's exact test, $p = 0.2262$; Supplementary Fig. S3e). Thus, we propose that the presence of the *MED12* mutation is associated with the molecular subtype of MPT (7/21 for "Epithelial," 0/16 for "Fibrous"; Fisher's exact test, $p = 0.0124$). The patients with "Epithelial" tumours showed significantly smaller tumour size compared to that of the patients with "Fibrous" tumours (Fig. 3c). These molecular subtypes were also significantly associated with the degree of stromal overgrowth, another well-known prognostic factor for MPT [54] (Fig. 3d). Although these molecular subtypes were markedly associated with histologic stromal overgrowth patterns, the histologic and molecular classifications showed similar metastasis patterns in both groups (Supplementary Fig. S3f and Supplementary Table S2).

Cancerous mesenchymal component and microenvironmental epithelial component of MPT

Among the nine MPT tumours used above, we obtained fresh tumour tissue from one patient ("mpt-07") during the curative surgery and develop a PDX model using NSG mice ("MX-99"). Previous studies have shown that murine cells replace the microenvironmental cells around human tumours after patient-derived tumours are engrafted into mice [55, 56]. Accordingly, we discriminated between the transcriptome of the tumour and the

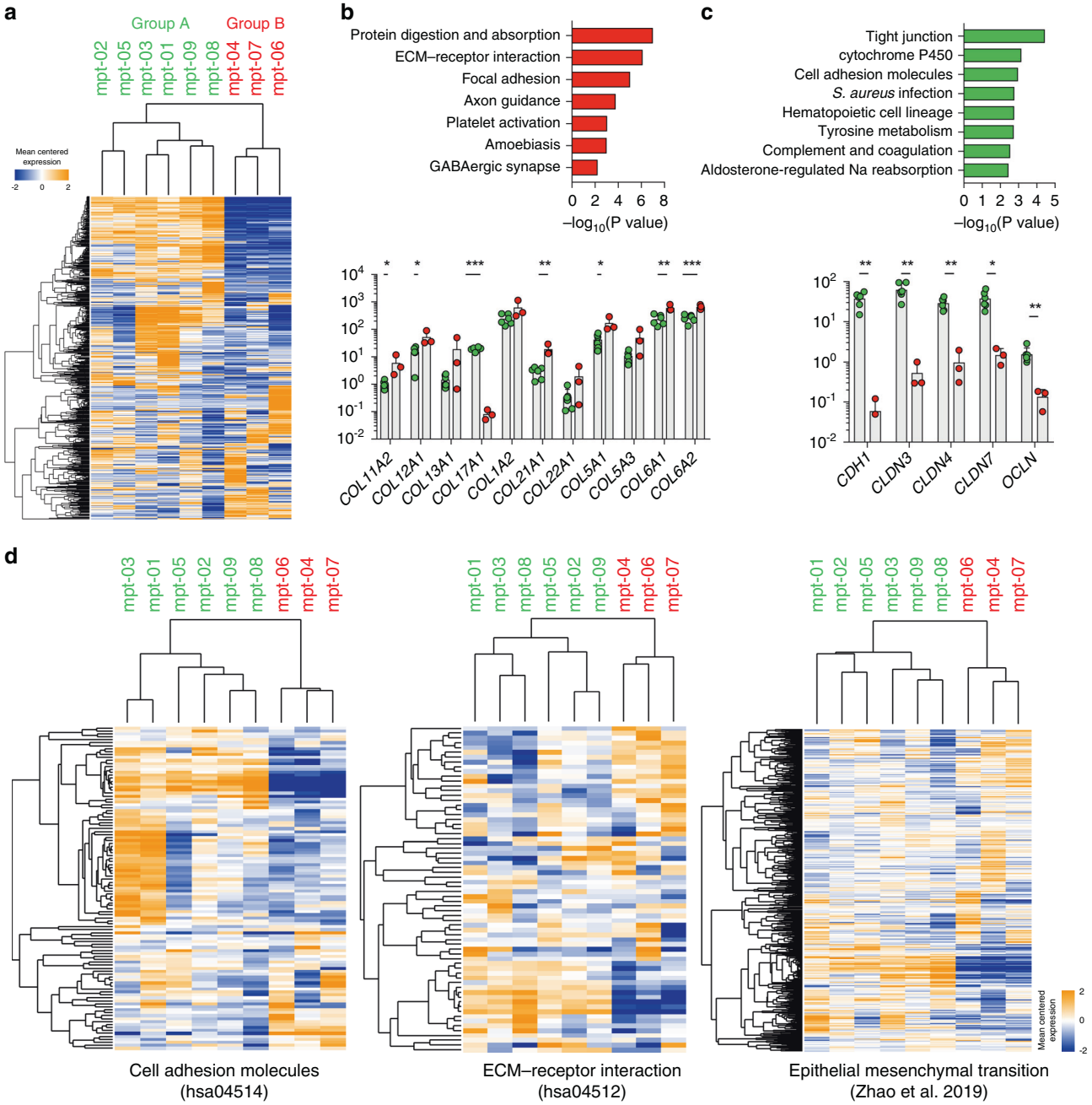


Fig. 2 Two molecular subtypes of malignant phyllodes tumours (MPTs). a Unsupervised clustering of nine MPT using expression levels of the top 2000 variable genes except for genes with low expression levels. **b** Significant pathways enriched with upregulated genes in "Group B" (top) and mRNA expression levels (FPKM) of various collagens (bottom) (Mann–Whitney test). **c** Significant pathways enriched with upregulated genes in "Group A" (top) and mRNA expression levels (FPKM) of epithelial cell markers such as cadherin, claudins, and occludin (bottom) (Mann–Whitney test). **d** Heatmaps showing the clustering of subtypes based on genes involved in cell adhesion molecules, ECM–receptor interactions, and EMT. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

microenvironment using species-specific genome sequences [12, 57]. Notably, when compared to those of PDX tumours derived from IDC, which are of epithelial origin, the RNA-seq data of PDX tumours derived from MPT tumours had a significantly higher proportion of mouse mRNA reads (Fig. 4a). Furthermore, when MPT PDX tumours were transplanted into multiple NSG mice and tumour tissues were collected at various tumour sizes, we observed that the proportion of mouse mRNA reads increased along with tumour size (Fig. 4b). These findings indicated that murine cells within the mouse microenvironment increase in number as the tumour grows. Histological examination of the MPT PDX tumours at different tumour sizes revealed that the mesenchymal component constitutes most of the tumour area during the early period, while the epithelial components gradually increased as the tumours grew (Fig. 4c, d).

The mouse reads obtained from MPT PDX tumours showed distinct gene expression patterns compared to those of normal mouse mammary fat pads, including upregulation of the cell cycle-related gene sets (Supplementary Fig. S4). To determine whether the epithelial-like morphology of separated murine cells was supported by gene expression profiles, we compared the expression levels of well-known epithelial and mesenchymal

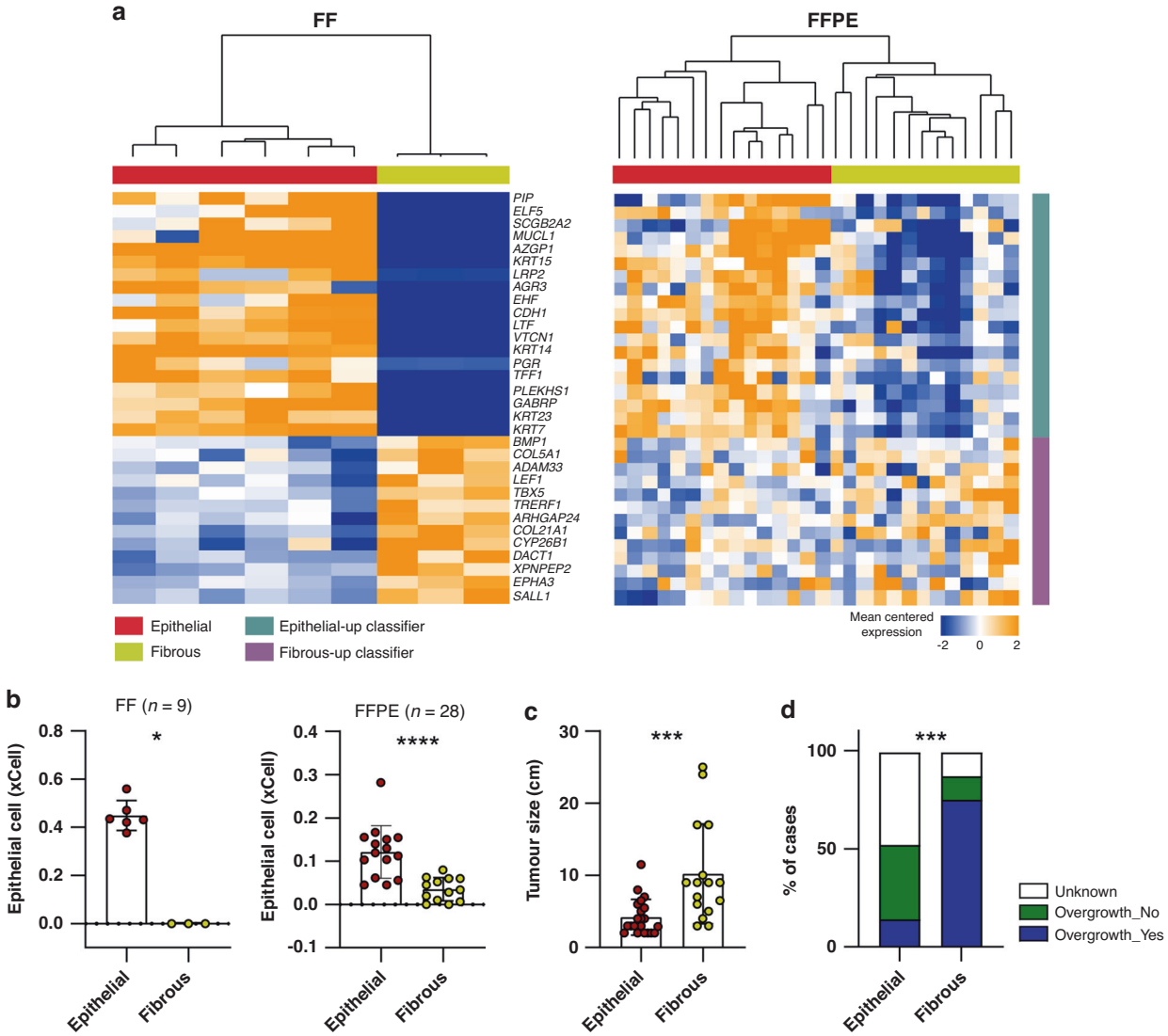
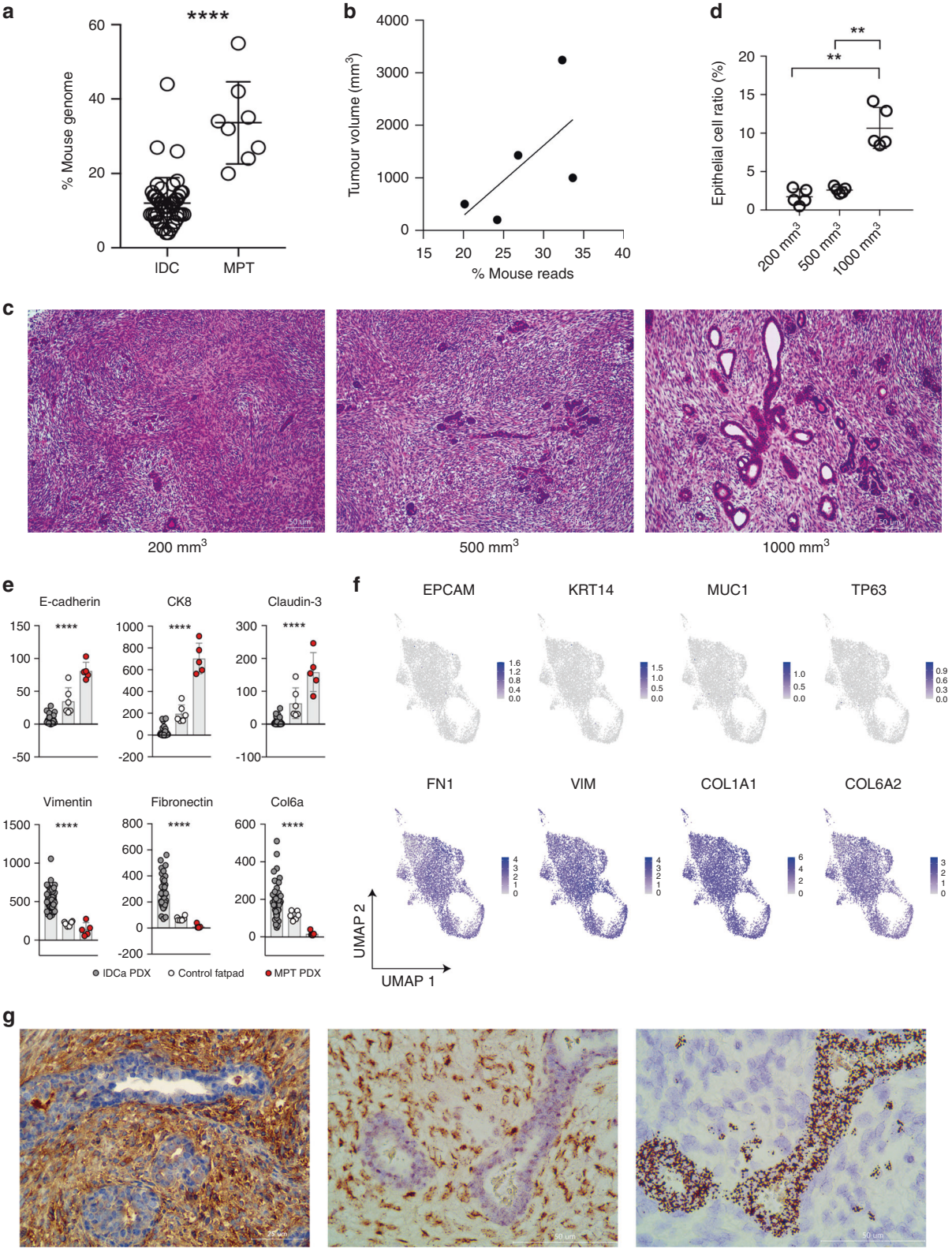


Fig. 3 Clinical implications of molecular subtypes of malignant phyllodes tumours (MPTs). a Hierarchical clustering of MPT FF (left) or FFPE (right) samples using the expression levels of the 32-classifier genes. **b** Comparison of the xCell scores of epithelial cells between the two groups in MPT FF (left) or FFPE (right) tissues (Mann–Whitney test). **c** Comparison of tumour size according to the molecular subtypes in MPT FF and FFPE tissues ("Epithelial", $n = 21$; "Fibrous", $n = 16$; Mann–Whitney test). **d** Comparison of degree of stromal overgrowth according to the molecular subtypes in MPT FF and FFPE tissues ("Epithelial", $n = 21$; "Fibrous", $n = 16$; Chi-square test). *$p < 0.05$, ***$p < 0.001$, ****$p < 0.0001$.

genes. When compared with the mouse reads from PDX tumours derived from IDC and normal mouse fat pads, mouse reads from the MPT PDX tumours showed upregulation of epithelial markers and downregulation of mesenchymal markers (Fig. 4e).

Thereafter, we performed scRNA-seq of the MPT PDX to confirm the identity of human MPT cells at single-cell resolution (Supplementary Fig. S5a). As shown in Fig. 4f, the expression of epithelial genes was low, whereas the expression of fibroblast genes was high in the cancerous human cells of the MPT PDX. We further isolated human and mouse cells from MPT PDX tumour tissues using the magnetic cell separation technique (Supplementary Fig. S5b) [58]. The isolated human tumour cells displayed a spindle-shaped, fibroblast-like appearance, while murine cells formed adherent structures similar to epithelial cells (Supplementary Fig S5c). To determine the origin of the cells more directly, we stained the MPT PDX tumour with a human-specific antibody and observed that only mesenchymal cells were positively stained with an anti-human HLA class 1 antibody (Fig. 4g). Furthermore, the epithelial cells stained positive for mouse-specific centromeric probes, while mesenchymal cells stained positive for human-specific probes by the species-specific FISH experiment [59] (Fig. 4g and Supplementary Fig. S5d). In addition, we developed a PDX model ("PDX78") using tissues obtained from another MPT patient and performed bulk RNA-seq analysis. Unlike the previous MPT PDX ("MX-99"), we did not find conclusive evidence showing an increase of the murine microenvironmental component in this MPT PDX ("PDX78") (Supplementary Fig. S5e). However, we confirmed that the expression of fibroblast genes, not epithelial genes, was also enriched in cancerous human cells of the "PDX78" as in the "MX-99" (Supplementary Fig. S5f).

Collectively, our results indicated that the mesenchymal cells of the MPT, and not the epithelial cells, have the capacity to survive and form PDX tumours within the mouse microenvironment. Furthermore, the murine epithelial component increased as it formed glandular structures within PDX tumours. These data suggested that, unlike carcinomas of epithelial origin, MPT consist of cancerous cells that have fibroblast-like features and a tumour microenvironment with proliferating epithelial cells.



Targeting mesenchymal component provides potential therapeutic approaches in MPT

Microscopic examination showed that the PDX tumour reproduced the histologic features of the patient's original tumour (Fig. 5a). The PDX tumour showed genomic and transcriptomic profiles similar to those of the primary tumour (Fig. 5b and Supplementary Fig. S6). The primary and PDX tumours had somatic mutations in *PDGFRB* (NM_002609: p.N666K) (Fig. 5b, c). *PDGFRB* N666K has been reported as an oncogenic mutation in rare mesenchymal disorders, infantile myofibromatosis, and cells that harbour the mutation respond to imatinib mesylate [60]. Recently, this mutation was also found in phyllodes tumours and

Fig. 4 Role of epithelial cells and stromal cells in malignant phyllodes tumours (MPTs). a Proportion of RNA-seq reads originating from murine cells in PDX derived from invasive ductal carcinoma (IDC) and MPT (Mann–Whitney test). **b** Correlation between MPT tumour volume and the proportion of RNA-seq reads originating from murine cells in the MPT PDX**. c** Haematoxylin and eosin staining results of MPT PDX tumour resected at volumes 200 mm^3 (left), 500 mm^3 (middle), and 1000 mm^3 (right). **d** Proportion of epithelial cells according to tumour volume (Mann–Whitney test). **e** Expression levels of epithelial cell markers (top line) and mesenchymal stromal cell markers (bottom line) based on mouse transcriptome in normal mouse fat pads and PDX derived from MPT and IDC (Kruskal–Wallis test). **f** UMAP embeddings of epithelial markers expressing grafted human tumour cells (top) and fibroblast markers expressing grafted human tumour cells (bottom) ($n = 8105$). **g** Identification of human mesenchymal cells and mouse epithelial cells by anti-human HLA class 1 staining (left), anti-FITC staining after human-specific centromeric FISH (middle), and mouse-specific centromeric FISH (right). $^{**}p < 0.01$, $^{****}p < 0.0001$.

was reported to respond to pazopanib, which is primarily used for soft tissue sarcomas [61]. *PDGFRB* is typically expressed in mesenchymal cells and not in epithelial cells [62]. Examination of RNA-seq data further revealed that when compared to IDC or normal breast tissues, *PDGFRB* mRNA expression was significantly upregulated in MPT samples, suggesting that *PDGFRB* levels may correlate with the degree of mesenchymal features (Fig. 5d and Supplementary Fig. S3c).

The "MX-99" tumour also harboured somatic mutation in *PIK3R1* (NM_181504: p.Y2fs), which is a regulatory subunit for PIK3CA with tumour-suppressor properties [63]. Somatic mutations in *PIK3R1* are often associated with increased PI3K/Akt signalling in cancer [64]. Our data also showed that MPT tissues had significant dysregulation of genes involved in the PI3K/Akt/mTOR pathway compared with IDC tissues (Fig. 5e). Additionally, *IGF1* and *IGF2*, which activate PI3K signalling in tumours in an autocrine manner, were also significantly upregulated in MPT tumours compared to IDC (Fig. 5f).

To test whether the genetic characteristics of the PDX model can be exploited to develop therapeutic approaches, we treated the PDX model with imatinib mesylate and PKI-587 [65] to target the PDGFR and PI3K/mTOR pathways, respectively. Both targeted agents significantly suppressed in vivo tumour growth of MPT in the PDX models, with higher efficacy for PKI-587 (Fig. 5g). Treatment with PKI-587 and imatinib mesylate effectively inhibited the phosphorylation of Akt and mTOR in PDX tumours, and imatinib inhibited PDGFRB phosphorylation (Fig. 5h, i). We additionally performed scRNA-seq analysis of drug-treated PDXs to identify changes in tumour cells upon drug treatment. Before comparative analysis between vehicle- and drug-treated PDXs, we first confirmed that human tumour cells in vehicle- and drug-treated MPT PDX had high expression of fibroblast genes, which is consistent with previous findings (Supplementary Fig. S7a). Through comparative analysis, we observed a decrease in the activity of the PI3K pathway and a decrease in the proportion of proliferating cells in the tumour cell population in drug-treated MPT PDX when compared with vehicle-treated MPT PDX (Supplementary Fig. S7b, c). These data demonstrated the potential efficacy of targeted approaches for MPT treatment.

DISCUSSION

In the present study, we characterised the genomic and transcriptomic features of a rare breast malignancy, MPT. The molecular features of MPT presented here provide novel insights into the pathogenesis of this biphasic neoplasm showing both mesenchymal and epithelial components. Furthermore, by using a PDX model derived from the MPT, we demonstrated the possibility of driver and microenvironment roles for each component. Finally, we showed that targeting the PDGFR and PI3K/mTOR pathways can be an effective treatment option for patients with MPT.

Previous studies of gene expression profiles of phyllodes tumours have mostly focused on the molecular characterisation of MPT within a wide range of fibroepithelial neoplasm [6–8]. For example, Vidal et al. [66] profiled gene expression data of 75 fibroepithelial tumours, including 11 MPT, and demonstrated that the dysregulation of epithelial- or luminal-related genes such as *CLDN3*, *CLDN7*, or apoptosis- and angiogenesis-related genes such as *VIM* or *PIK3CA* can be key molecular characteristics of MPT. In contrast, the present study included only pathologically proven MPT to identify the molecular features that can classify malignant tumours with different clinical characteristics. Our data suggested that the expression patterns of genes involved in ECM interactions, EMT, and cell adhesion could classify MPT into two subtypes. The fact that the two subtypes of MPT are characterised by genes related to epithelial and mesenchymal differentiation indicated that the biphasic pathological features of fibroepithelial neoplasms are also reflected at the molecular level. Moreover, we showed that the molecular subtype of MPT is significantly associated with the degree of epithelial cell proportion, tumour size, and stromal overgrowth. We further inferred that the presence of mutations in *MED12* may be related to the molecular subtype of MPT. Notably, the *MED12* mutation, found in the stroma [10, 67], appears frequently in the group in which the epithelial component is relatively high. As this study had a small number of samples, further studies using sufficient MPT samples to prove this will need to be carried out in the future.

Sawhney et al. previously suggested that stromal growth in fibroepithelial tumours may be affected by the epithelium, showing that stromal mitosis is more prevalent in stroma close to the epithelium [68]. Tse et al. suggested the stromal-epithelial interaction and the crucial role of the epithelium in tumour progression of phyllodes tumours, showing that epithelial ER expression was inversely related to stromal mitotic count [69]. Since then, few studies have been conducted on the relationship between the epithelial and stromal components in fibroepithelial neoplasms. In most previous studies, each component was separated and analysed using methods such as laser capture microdissection to unveil the molecular characteristics of the epithelial and stromal components [10, 67]. In the present study, we showed the molecular characteristics and roles of each of the epithelial and stromal components through bioinformatic deconvolution using bulk RNA-seq data for the MPT PDX model, suggesting that this method can enhance the understanding of the pathogenesis of fibroepithelial neoplasms consisting of two components.

The genomic and transcriptomic findings in our study identified two druggable pathways in MPT: the PDGFR and PI3K/mTOR pathways. PDGFRα and PDGFRβ proteins are upregulated in stromal cells of human cancers [70] and are expressed in 10–70% of borderline or MPTs [71]. The stromal activity of PDGFR is associated with histologic grades and treatment outcomes of phyllodes tumours [52]. Moreover, the PI3K/mTOR pathway has been identified as a critical regulatory pathway in MPT [72, 73]. We showed that pharmacological inhibition of the PDGFR and PI3K/mTOR pathways with imatinib mesylate and PKI-587 effectively suppressed tumour growth in vivo. Our study shows targeted approaches that are viable therapeutic tools for the treatment of MPT. The proposed therapeutic approach should be tested in future clinical trials.

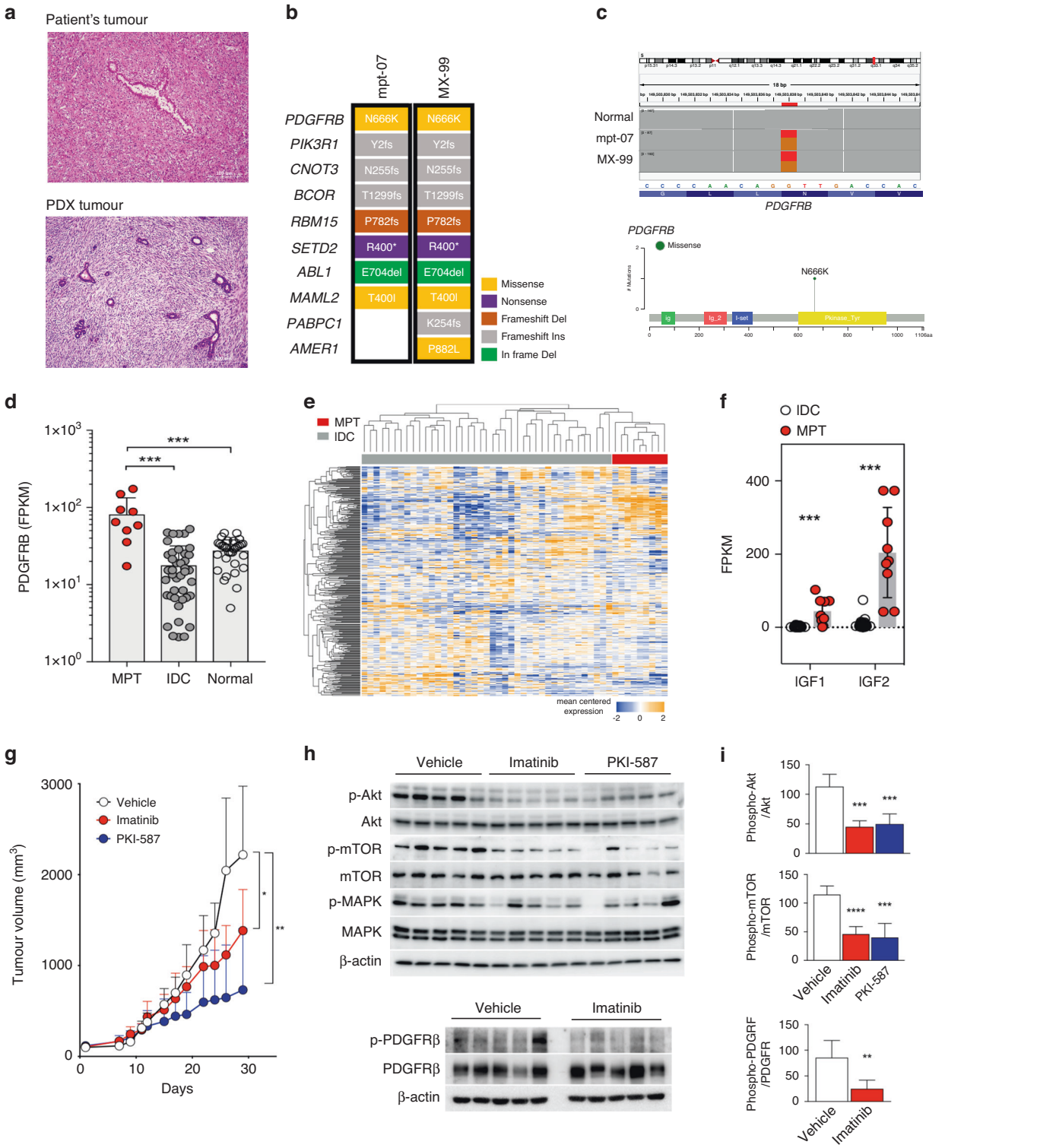


Fig. 5 Discovery of potential therapeutic target of malignant phyllodes tumours (MPTs). **a** Hematoxylin and eosin stained picture of primary tumour "mpt-07" of the patient with MPT (top) and MPT PDX tumour "MX-99" (bottom). **b** Somatic mutations in cancer-related genes in primary tumours (mpt-07) and matched PDX tumours (MX-99). **c** Visualisation in Integrative Genomics Viewer (IGV) shows p.N666K somatic mutation of *PDGFRB* in the primary and PDX tumours (top), and the lollipop plot shows the p.N666K mutation in the *PDGFRB* kinase domain (bottom). **d** *PDGFRB* mRNA expression levels in normal breast tissues, invasive ductal carcinoma (IDC) tissues, and MPT (Mann–Whitney test). **e** Heatmap showing the expression patterns of genes involved in the PI3K/Akt/mTOR pathway in MPT and IDC. Genes with an average FPKM across all samples ≥1 were selected among genes in the PI3K-Akt signalling pathway (KEGG ID: hsa04151) and mTOR signalling pathway (KEGG ID: hsa04150). **f** mRNA expression levels of *IGF1* and *IGF2* in MPT and IDC (Mann–Whitney test). **g** Tumour growth in the "MX-99" xenograft model treated by vehicle, imatinib, and PKI-587 (Wilcoxon test). **h** Western blot analysis of downstream signalling pathway molecules in xenograft tumours treated with vehicle, imatinib, and PKI-587. **i** Quantitative analysis of western blot results (Mann–Whitney test). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

DATA AVAILABILITY

The data sets used and/or analysed during the current study are available from the corresponding author on reasonable request.

REFERENCES

1. Yang X, Kandil D, Cosar EF, Khan A. Fibroepithelial tumors of the breast: pathologic and immunohistochemical features and molecular mechanisms. Arch Pathol Lab Med. 2014;138:25–36.
2. Spitaleri G, Toesca A, Botteri E, Bottiglieri L, Rotmensz N, Boselli S, et al. Breast phyllodes tumor: a review of literature and a single center retrospective series analysis. Crit Rev Oncol Hematol. 2013;88:427–36.
3. Gnerlich JL, Williams RT, Yao K, Jaskowiak N, Kulkarni SA. Utilization of radiotherapy for malignant phyllodes tumors: analysis of the National Cancer Data Base, 1998-2009. Ann Surg Oncol. 2014;21:1222–30.
4. Tan BY, Acs G, Apple SK, Badve S, Bleiweiss IJ, Brogi E, et al. Phyllodes tumours of the breast: a consensus review. Histopathology. 2016;68:5–21.
5. Karim RZ, O'Toole SA, Scolyer RA, Cooper CL, Chan B, Selinger C, et al. Recent insights into the molecular pathogenesis of mammary phyllodes tumours. J Clin Pathol. 2013;66:496–505.
6. Lae M, Vincent-Salomon A, Savignoni A, Huon I, Freneaux P, Sigal-Zafrani B, et al. Phyllodes tumors of the breast segregate in two groups according to genetic criteria. Mod Pathol. 2007;20:435–44.
7. Jones AM, Mitter R, Poulsom R, Gillett C, Hanby AM, Tomlinson IP, et al. mRNA expression profiling of phyllodes tumours of the breast: identification of genes important in the development of borderline and malignant phyllodes tumours. J Pathol. 2008;216:408–17.
8. Ang MK, Ooi AS, Thike AA, Tan P, Zhang Z, Dykema K, et al. Molecular classification of breast phyllodes tumors: validation of the histologic grading scheme and insights into malignant progression. Breast Cancer Res Treat. 2011;129:319–29.
9. Cani AK, Hovelson DH, McDaniel AS, Sadis S, Haller MJ, Yadati V, et al. Next-gen sequencing exposes frequent MED12 mutations and actionable therapeutic targets in phyllodes tumors. Mol Cancer Res. 2015;13:613–9.
10. Tan J, Ong CK, Lim WK, Ng CC, Thike AA, Ng LM, et al. Genomic landscapes of breast fibroepithelial tumors. Nat Genet. 2015;47:1341–5.
11. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26:589–95.
12. Cho SY, Chae J, Na D, Kang W, Lee A, Min S, et al. Unstable genome and transcriptome dynamics during tumor metastasis contribute to therapeutic heterogeneity in colorectal cancers. Clin Cancer Res. 2019;25:2821–34.
13. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43:11.10.1–.33.
14. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.
15. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31:213–9.
16. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.
17. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016;17:31.
18. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2019;47:D941–D7.
19. Gulhan DC, Lee JJ, Melloni GEM, Cortes-Ciriano I, Park PJ. Detecting the mutational signature of homologous recombination deficiency in clinical samples. Nat Genet. 2019;51:912–9.
20. Kwak SH, Chae J, Choi S, Kim MJ, Choi M, Chae JH, et al. Findings of a 1303 Korean whole-exome sequencing study. Exp Mol Med. 2017;49:e356.
21. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6:pl1.
22. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2:401–4.
23. Ramos AH, Lichtenstein L, Gupta M, Lawrence MS, Pugh TJ, Saksena G, et al. Oncotator: cancer variant annotation tool. Hum Mutat. 2015;36:E2423–9.
24. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat Rev Cancer. 2018;18:696–705.
25. Magi A, Tattini L, Cifola I, D'Aurizio R, Benelli M, Mangano E, et al. EXCAVATOR: detecting copy number variants from whole-exome sequencing data. Genome Biol. 2013;14:R120.
26. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. PLoS Comput Biol. 2016;12:e1004873.
27. Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP, et al. Copy number variation detection and genotyping from exome sequence data. Genome Res. 2012;22:1525–32.
28. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
29. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31:166–9.
30. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.
31. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
32. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009;25:1091–3.
33. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28:27–30.
34. de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. Bioinformatics. 2004;20:1453–4.
35. Saldanha AJ. Java Treeview—extensible visualization of microarray data. Bioinformatics. 2004;20:3246–8.
36. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017;8:14049.
37. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. Cell Syst. 2019;8:329–37.e4.
38. Young MD, Behjati S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. Gigascience. 2020;9:giaa151.
39. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. 2015;33:495–502.
40. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16:1289–96.
41. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500:415–21.
42. Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat Genet. 2013;45:970.
43. Piscuoglio S, Ng CKY, Murray M, Burke KA, Edelweiss M, Geyer FC, et al. Massively parallel sequencing of phyllodes tumours of the breast reveals actionable mutations, and TERT promoter hotspot mutations and TERT gene amplification as likely drivers of progression. J Pathol. 2016;238:508–18.
44. Yoon N, Bae GE, Kang SY, Choi MS, Hwang HW, Kim SW, et al. Frequency of MED12 mutations in phyllodes tumors: inverse correlation with histologic grade. Genes Chromosomes Cancer. 2016;55:495–504.
45. Yoshida M, Sekine S, Ogawa R, Yoshida H, Maeshima A, Kanai Y, et al. Frequent MED12 mutations in phyllodes tumours of the breast. Br J Cancer. 2015;112:1703–8.
46. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.
47. Lu P, Weaver VM, Werb Z. The extracellular matrix: a dynamic niche in cancer progression. J Cell Biol. 2012;196:395–406.
48. Polyak K, Weinberg RA. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. Nat Rev Cancer. 2009;9:265–73.
49. Ikenouchi J, Matsuda M, Furuse M, Tsukita S. Regulation of tight junctions during the epithelium-mesenchyme transition: direct repression of the gene expression of claudins/occludin by Snail. J Cell Sci. 2003;116(Pt 10):1959–67.
50. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. 2010;12:R68.
51. Zhao M, Liu Y, Zheng C, Qu H. dbEMT 2.0: An updated database for epithelial-mesenchymal transition genes with experimentally verified information and precalculated regulation information for cancer metastasis. J Genet Genomics. 2019;46:595–7.
52. Kim HM, Lee YK, Koo JS. Expression of CAF-related proteins is associated with histologic grade of breast phyllodes tumor. Dis Markers. 2016;2016:4218989.
53. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. 2017;18:220.
54. Chen WH, Cheng SP, Tzen CY, Yang TL, Jeng KS, Liu CL, et al. Surgical treatment of phyllodes tumors of the breast: retrospective review of 172 cases. J Surg Oncol. 2005;91:185–94.

55. Wang T, Lu R, Kapur P, Jaiswal BS, Hannan R, Zhang Z, et al. An empirical approach leveraging tumorgrafts to dissect the tumor microenvironment in renal cell carcinoma identifies missing link to prognostic inflammatory factors. Cancer Discov. 2018;8:1142–55.
56. Blomme A, Van Simaeys G, Doumont G, Costanza B, Bellier J, Otaka Y, et al. Murine stroma adopts a human-like metabolic phenotype in the PDX model of colorectal cancer and liver metastases. Oncogene. 2018;37:1237–50.
57. Callari M, Batra AS, Batra RN, Sammut SJ, Greenwood W, Clifford H, et al. Computational approach to discriminate human and mouse sequences in patient-derived tumour xenografts. BMC Genomics. 2018;19:19.
58. Agorku DJ, Tomiuk S, Klingner K, Wild S, Ruberg S, Zatrieb L, et al. Depletion of mouse cells from human tumor xenografts significantly improves downstream analysis of target cells. J Vis Exp. 2016:54259.
59. Bruna A, Rueda OM, Greenwood W, Batra AS, Callari M, Batra RN, et al. A biobank of breast cancer explants with preserved intra-tumor heterogeneity to screen anticancer compounds. Cell. 2016;167:260.e22–74.e22.
60. Arts FA, Chand D, Pecquet C, Velghe AI, Constantinescu S, Hallberg B, et al. PDGFRB mutants found in patients with familial infantile myofibromatosis or overgrowth syndrome are oncogenic and sensitive to imatinib. Oncogene. 2016;35:3239–48.
61. Ochi M, Miyamoto S, Nakao K, Furuhata Y, Tabei Y, Ichi S, et al. Significant antitumor activity of pazopanib in a patient with PDGFRB-mutated metastatic phyllodes tumor: a case report. SN Compr Clin Med. 2021;3:901–5.
62. Paulsson J, Sjoblom T, Micke P, Ponten F, Landberg G, Heldin CH, et al. Prognostic significance of stromal platelet-derived growth factor beta-receptor expression in human breast cancer. Am J Pathol. 2009;175:334–41.
63. Taniguchi CM, Winnay J, Kondo T, Bronson RT, Guimaraes AR, Aleman JO, et al. The phosphoinositide 3-kinase regulatory subunit p85alpha can exert tumor suppressor properties through negative regulation of growth factor signaling. Cancer Res. 2010;70:5305–15.
64. Engelman JA. Targeting PI3K signalling in cancer: opportunities, challenges and limitations. Nat Rev Cancer. 2009;9:550–62.
65. Shapiro GI, Bell-McGuinn KM, Molina JR, Bendell J, Spicer J, Kwak EL, et al. First-in-human study of PF-05212384 (PKI-587), a Small-molecule, intravenous, dual inhibitor of PI3K and mTOR in patients with advanced cancer. Clin Cancer Res. 2015;21:1888–95.
66. Vidal M, Peg V, Galvan P, Tres A, Cortes J, Ramon Y Cajal S, et al. Gene expression-based classifications of fibroadenomas and phyllodes tumours of the breast. Mol Oncol. 2015;9:1081–90.
67. Lim WK, Ong CK, Tan J, Thike AA, Ng CC, Rajasegaran V, et al. Exome sequencing identifies highly recurrent MED12 somatic mutations in breast fibroadenoma. Nat Genet. 2014;46:877–80.
68. Sawhney N, Garrahan N, Douglas-Jones AG, Williams ED. Epithelial—stromal interactions in tumors. A morphologic study of fibroepithelial tumors of the breast. Cancer. 1992;70:2115–20.
69. Tse GM, Lee CS, Kung FY, Scolyer RA, Law BK, Lau TS, et al. Hormonal receptors expression in epithelial cells of mammary phyllodes tumors correlates with pathologic grade of the tumor: a multicenter study of 143 cases. Am J Clin Pathol. 2002;118:522–6.
70. Paulsson J, Ehnman M, Ostman A. PDGF receptors in tumor biology: prognostic and predictive potential. Future Oncol. 2014;10:1695–708.
71. Feakins RM, Wells CA, Young KA, Sheaff MT. Platelet-derived growth factor expression in phyllodes tumors and fibroadenomas of the breast. Hum Pathol. 2000;31:1214–22.
72. Liu SY, Joseph NM, Ravindranathan A, Stohr BA, Greenland NY, Vohra P, et al. Genomic profiling of malignant phyllodes tumors reveals aberrations in FGFR1 and PI-3 kinase/RAS signaling pathways and provides insights into intratumoral heterogeneity. Mod Pathol. 2016;29:1012–27.
73. Gatalica Z, Vranic S, Ghazalpour A, Xiu J, Ocal IT, McGill J, et al. Multiplatform molecular profiling identifies potentially targetable biomarkers in malignant phyllodes tumors of the breast. Oncotarget. 2016;7:1707–16.

AUTHOR CONTRIBUTIONS

CL, H-GM, and J-IK conceived and designed the study. E-SL, T-KY, D-YN, and H-GM provided patient samples and clinical data. E-SL and T-KY managed and analysed samples and clinical data. DN, WK, JK, and CL provided materials and resources related to mice, including the patient-derived xenograft (PDX) model. JY, JC, and DL performed bioinformatic analyses. WH and WK conducted experiments. WL and JH provided substantial assistance in bioinformatic analyses. JK, BSH, and H-YS provided substantial assistance with the experiments. JY, WH, E-SL, DN, JC, H-GM, and J-IK interpreted and improved the results. D-YN and CL reviewed scientific and clinical data. JY, WH, E-SL, H-GM, and J-IK wrote the initial manuscript. H-GM and J-IK supervised the study.

FUNDING

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (Grant nos. 2019R1A2C2005277, 2020R1A6A1A03047972, 2018R1D1A1B07045381), and by a grant from the Korea Health Industry Development Institute (KHIDI) (Grant nos. HI22C0497, HA15C0011) and a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) (Grant no. HI14C1277) funded by the Ministry of Health and Welfare, Republic of Korea. CL was a distinguished Ewha Womans University Professor supported in part by the Ewha Womans University Research grant of 2018-2019. This study was supported in part by operational funds from The First Affiliated Hospital of Xi'an Jiaotong University.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The acquisition and use of specimens and clinical and pathologic data from breast cancer patients for this study were approved by the institutional IRB (IRB no. 1509-032-702 and IRB no. 1402-054-555). The details are included in the "Methods" section. Written informed consent was obtained from all study participants. All procedures were performed in accordance with the Declaration of Helsinki. The mice were cared for according to institutional guidelines for animal care. All animal experiments were approved by the Institutional Animal Care and Use Committee (IACUC) of Seoul National University (no. 14-0016-C0A0).

CONSENT FOR PUBLICATION

Not applicable.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41416-022-02064-2.

Correspondence and requests for materials should be addressed to Hyeong-Gon Moon or Jong-Il Kim.

Reprints and permission information is available at http://www.nature.com/reprints

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

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

© The Author(s) 2022