Lawrence Berkeley National Laboratory

Title

NCAPH drives breast cancer progression and identifies a gene signature that predicts luminal a tumour recurrence.

Permalink

https://escholarship.org/uc/item/29d362ks

Journal

Clinical and Translational Medicine, 14(2)

Authors

Mendiburu-Eliçabe, Marina

García-Sancha, Natalia

Corchado-Cobos, Roberto

et al.

Publication Date

2024-02-01

DOI

10.1002/ctm2.1554

Peer reviewed

CLINICAL AND TRANSLATIONAL MEDICINE

WILEY

RESEARCH ARTICLE

NCAPH drives breast cancer progression and identifies a gene signature that predicts luminal a tumour recurrence

Marina Mendiburu-Eliçabe[1,2]　|　Natalia García-Sancha[1,2]　|
Roberto Corchado-Cobos[1,2]　|　Angélica Martínez-López[3,4]　|　Hang Chang[5,6]　|
Jian Hua Mao[5,6]　|　Adrián Blanco-Gómez[1,2]　|　Ana García-Casas[3,4]　|
Andrés Castellanos-Martín[1,2]　|　Nélida Salvador[3,4]　|　Alejandro Jiménez-Navas[1,2]　|
Manuel Jesús Pérez-Baena[1,2]　|　Manuel Adolfo Sánchez-Martín[7,8]　|
María Del Mar Abad-Hernández[2,9,10]　|　Sofía Del Carmen[2,9,10]　|
Juncal Claros-Ampuero[1,2,11]　|　Juan Jesús Cruz-Hernández[1,2,7,11]　|
César Augusto Rodríguez-Sánchez[1,2,7,11]　|　María Begoña García-Cenador[2,12]　|
Francisco Javier García-Criado[2,12]　|　Rodrigo Santamaría Vicente[13]　|
Sonia Castillo-Lluva[3,4]　|　Jesús Pérez-Losada[1,2]

[1]Instituto de Biología Molecular y Celular del Cáncer (IBMCC-CIC), Universidad de Salamanca/CSIC, Salamanca, Spain

[2]Biosanitary Research Institute of Salamanca (IBSAL), Salamanca, Spain

[3]Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Químicas, Universidad Complutense, Madrid, Spain

[4]San Carlos Health Research Institute (IdISSC), Madrid, Spain

[5]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory (LBNL), Berkeley, California, USA

[6]Berkeley Biomedical Data Science Center, Lawrence Berkeley National Laboratory (LBNL), Berkeley, California, USA
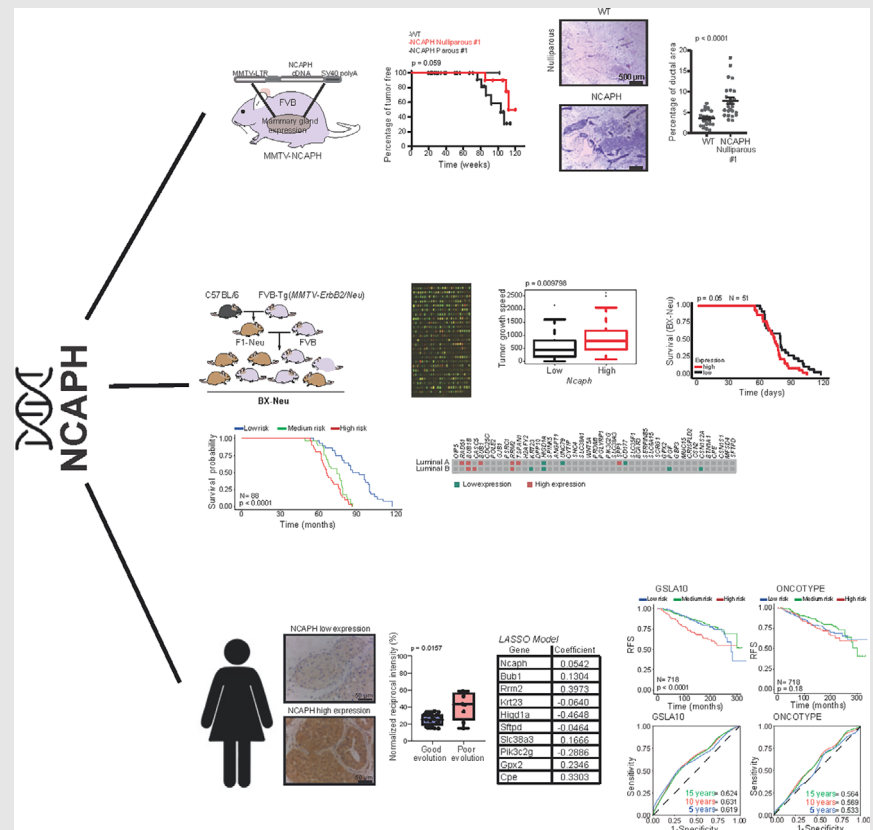
[7]Departamento de Medicina, Universidad de Salamanca, Salamanca, Spain

[8]Servicio de Transgénesis, Plataforma Nucleus, Universidad de Salamanca, Salamanca, Spain

[9]Departamento de Anatomía Patológica, Universidad de Salamanca, Salamanca, Spain

[10]Servicio de Anatomía Patológica, Hospital Universitario de Salamanca, Salamanca, Spain

Graphical Abstract

[11]Servicio de Oncología, Hospital Universitario de Salamanca, Salamanca, Spain

[12]Departamento de Cirugía, Universidad de Salamanca, Salamanca, Spain

[13]Departamento de Informática y Automática, Universidad de Salamanca, Salamanca, España

Correspondence

Jesús Pérez-Losada, Instituto de Biología Molecular y Celular del Cáncer (IBMCC-CIC), Universidad de Salamanca/CSIC, Salamanca, Spain.
Email: jperezlosada@usal.es

Sonia Castillo-Lluva, Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Químicas, Universidad Complutense, Madrid, Spain.
Email: sonica01@ucm.es

- NCAPH significantly influences luminal A breast cancer progression and affects tumour behaviour and treatment response.
- GSLA10, a gene signature associated with NCAPH, shows promise in predicting luminal A tumour recurrence, demonstrating efficacy alongside existing models such as Oncotype DX.
- The study highlights the potential for tailored treatment approaches in luminal A breast cancer based on NCAPH and GSLA10 insights.

CLINICAL AND TRANSLATIONAL MEDICINE
WILEY

RESEARCH ARTICLE

NCAPH drives breast cancer progression and identifies a gene signature that predicts luminal a tumour recurrence

Marina Mendiburu-Eliçabe[1,2] | **Natalia García-Sancha**[1,2] |
Roberto Corchado-Cobos[1,2] | **Angélica Martínez-López**[3,4] | **Hang Chang**[5,6] |
Jian Hua Mao[5,6] | **Adrián Blanco-Gómez**[1,2] | **Ana García-Casas**[3,4] |
Andrés Castellanos-Martín[1,2] | **Nélida Salvador**[3,4] | **Alejandro Jiménez-Navas**[1,2] |
Manuel Jesús Pérez-Baena[1,2] | **Manuel Adolfo Sánchez-Martín**[7,8] |
María Del Mar Abad-Hernández[2,9,10] | **Sofía Del Carmen**[2,9,10] |
Juncal Claros-Ampuero[1,2,11] | **Juan Jesús Cruz-Hernández**[1,2,7,11] |
César Augusto Rodríguez-Sánchez[1,2,7,11] | **María Begoña García-Cenador**[2,12] |
Francisco Javier García-Criado[2,12] | **Rodrigo Santamaría Vicente**[13] |
Sonia Castillo-Lluva[3,4] | **Jesús Pérez-Losada**[1,2]

[1]Instituto de Biología Molecular y Celular del Cáncer (IBMCC-CIC), Universidad de Salamanca/CSIC, Salamanca, Spain

[2]Biosanitary Research Institute of Salamanca (IBSAL), Salamanca, Spain

[3]Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Químicas, Universidad Complutense, Madrid, Spain

[4]San Carlos Health Research Institute (IdISSC), Madrid, Spain

[5]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory (LBNL), Berkeley, California, USA

[6]Berkeley Biomedical Data Science Center, Lawrence Berkeley National Laboratory (LBNL), Berkeley, California, USA

[7]Departamento de Medicina, Universidad de Salamanca, Salamanca, Spain

[8]Servicio de Transgénesis, Plataforma Nucleus, Universidad de Salamanca, Salamanca, Spain

[9]Departamento de Anatomía Patológica, Universidad de Salamanca, Salamanca, Spain

[10]Servicio de Anatomía Patológica, Hospital Universitario de Salamanca, Salamanca, Spain

[11]Servicio de Oncología, Hospital Universitario de Salamanca, Salamanca, Spain

[12]Departamento de Cirugía, Universidad de Salamanca, Salamanca, Spain

[13]Departamento de Informática y Automática, Universidad de Salamanca, Salamanca, España

Correspondence
Jesús Pérez-Losada, Instituto de Biología Molecular y Celular del Cáncer (IBMCC-CIC), Universidad de Salamanca/CSIC, Salamanca, Spain.

Abstract

Background: Luminal A tumours generally have a favourable prognosis but possess the highest 10-year recurrence risk among breast cancers. Additionally, a quarter of the recurrence cases occur within 5 years post-diagnosis.

Email: jperezlosada@usal.es

Sonia Castillo-Lluva, Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Químicas, Universidad Complutense, Madrid, Spain. Email: sonica01@ucm.es

Identifying such patients is crucial as long-term relapsers could benefit from extended hormone therapy, while early relapsers might require more aggressive treatment.

Methods: We conducted a study to explore non-structural chromosome maintenance condensin I complex subunit H's (NCAPH) role in luminal A breast cancer pathogenesis, both in vitro and in vivo, aiming to identify an intratumoural gene expression signature, with a focus on elevated NCAPH levels, as a potential marker for unfavourable progression. Our analysis included transgenic mouse models overexpressing NCAPH and a genetically diverse mouse cohort generated by backcrossing. A least absolute shrinkage and selection operator (LASSO) multivariate regression analysis was performed on transcripts associated with elevated intratumoural NCAPH levels.

Results: We found that NCAPH contributes to adverse luminal A breast cancer progression. The intratumoural gene expression signature associated with elevated NCAPH levels emerged as a potential risk identifier. Transgenic mice overexpressing NCAPH developed breast tumours with extended latency, and in Mouse Mammary Tumor Virus (MMTV)-$NCAPH^{ErbB2}$ double-transgenic mice, luminal tumours showed increased aggressiveness. High intratumoural Ncaph levels correlated with worse breast cancer outcome and subpar chemotherapy response. A 10-gene risk score, termed Gene Signature for Luminal A 10 (GSLA10), was derived from the LASSO analysis, correlating with adverse luminal A breast cancer progression.

Conclusions: The GSLA10 signature outperformed the Oncotype DX signature in discerning tumours with unfavourable outcomes, previously categorised as luminal A by Prediction Analysis of Microarray 50 (PAM50) across three independent human cohorts. This new signature holds promise for identifying luminal A tumour patients with adverse prognosis, aiding in the development of personalised treatment strategies to significantly improve patient outcomes.

KEYWORDS
breast cancer, genetic signature, LASSO, luminal A subtype, NCAPH, prognosis, relapse-free survival

1 | BACKGROUND

Breast cancer is currently the most frequently diagnosed tumour worldwide and the leading cause of cancer-related deaths in women.[1] In recent years, assays have been designed to evaluate the risk of breast cancer relapse based on the expression of multiple genes, thereby enabling the selection of the best treatment option.[2–8] Through these tests, intrinsic subtypes of breast cancer can be defined to enhance the ability to predict the outcome of this disease.[5,9–11]

One of the limitations of these signatures is that their long-term predictive ability has not yet been fully determined, particularly in luminal A tumours.[10,12] Luminal A tumours have the most favourable prognosis, with the highest disease-free survival (DFS) rate at 10 years; however, it should be noted that 25% of these tumours that relapse do it within the first 5 years, and the relapse risk notably triples between the 5th and 10th years.[13] Furthermore, luminal A tumours are most prone to recurrence beyond the 10-year follow-up mark.[14] Indeed, metastasis may appear many years after the diagnosis of luminal A breast cancer and even decades after the removal of the primary tumour.[15,16] Additionally, treated luminal A tumours that eventually relapse may evolve into aggressive metastatic cancers.[17] Therefore, it is paramount to identify luminal A tumours with a heightened relapse risk in the short to medium term, as these may require

a more aggressive therapeutic approach. Similarly, it is essential to identify patients at an elevated risk of long-term recurrence to enable meticulous monitoring and potentially extend their secondary chemoprevention hormone therapy.[14] Thus, there is a clear need to identify gene signatures for tumours already defined as luminal A that could distinguish those with potentially worse outcome in short-, medium- and long-term follow-up. Indeed, including genes that are involved in the pathogenesis of luminal A tumours with poor outcome is likely to be useful for obtaining signatures that may more accurately predict prognosis.[18] In this sense, luminal A tumours with a worse prognosis would exhibit higher genomic instability, including multiple modifications to the focal copy number[19] and alterations in genes involved in mitosis.[20]

NCAPH encodes a member of the Barr protein family that contributes to the condensin complex. Condensins are large protein complexes that assemble interphase chromatin into chromosomes and organise their segregation during mitosis and meiosis.[21] Two condensin complexes have been described,[22] of which the condensin II complex is predominantly located in the nucleus during interphase and binds to chromosomes during mitosis.[23–25] Conversely, the condensin I complex is located in the cytoplasm and binds to chromosomes after rupture of the nuclear membrane.[26] However, a small fraction of condensin I has been detected in the nucleus during interphase, where it helps regulate gene expression and chromosome condensation.[26] These two multiprotein complexes share two central subunits: the structural chromosome maintenance proteins SMC2 and SMC4. They also contain three non-structural chromosome maintenance (SMC) subunits: NCAPD2, NCAPG and non-SMC condensin I complex subunit H (NCAPH) in the condensin I complex and NCAPD3, NCAPG2 and NCAPH2 in the condensin II complex. The NCAPH and NCAPH2 subunits belong to the kleisin protein family,[27–29] and mutations in the latter have been implicated in genomic instability.[30]

Here, we demonstrate that NCAPH participates in the pathogenesis of breast cancer and in chemotherapy resistance in vivo and in vitro. MMTV-*NCAPH*[ErbB2] double-transgenic mice overexpressing NCAPH generated more aggressive breast tumours, and in a cohort of genetically heterogeneous transgenic mice generated by backcrossing, these tumours had a worse outcome and a poor response to chemotherapy. Moreover, using the least absolute shrinkage and selection operator (LASSO) multivariate regression model,[31] we identified a group of 10 genes associated with high intratumoural NCPAH levels from this cohort of mice that formed a signature capable of defining patients with a worse outcome of luminal A tumours, better defining those who could benefit from more personalised treatment.

2 | METHODS

2.1 | Patient samples and immunostaining

Human primary breast tumours were collected at the University Hospital of Salamanca, after the Hospital's Institutional Ethics Review Board approved the protocols for the collection and use of patient samples. Written informed consent was obtained from all patients to conduct the study on these tumour samples. NCAPH expression was evaluated in a retrospective study of 28 human tumour samples by immunohistochemistry (IHC), eight of which were from patients who had a poor outcome and developed liver metastases, while the other 20 had a good outcome. IHC was performed automatically using the Bond Polymer Refine Detection kit (BOND III: Leica, Biosystems, Leica Microsystems), probing the sections for 60 min with an antibody against NCAPH (HPA0030008, Sigma–Aldrich) diluted 1:100 and then counterstained with haematoxylin. Appropriate positive and negative controls were used.

Fiji Is Just ImageJ (FIJI) software was used to analyse NCAPH expression in human breast tumours.[32] FIJI, is an open-source, Java-based image processing platform. Tailored for biological image analysis, FIJI provides a comprehensive suite of tools for advanced processing and quantitative assessment, including densitometry. This software ensures precise and replicable analysis of microscopic images,[32] sampling seven fields for each tumour randomly with a Leica ICC50 HD camera at 40× magnification under the control of the Leica Application Suite V3.7 software. The mean intensity (MI) of the epithelial cells was obtained and the reciprocal intensity (RI) was calculated in each field by subtracting the epithelial MI from the background.[33] The normalised RI was obtained by dividing the RI by the background value, and the mean normalised RI of the seven fields was calculated for each patient. For more details on image analysis using FIJI software, see the Supporting Information.

2.2 | NCAPH transgenic mice

All mice were housed at the Animal Research Facility of the University of Salamanca and all procedures were approved by the Institutional Animal Care and Bioethical Committee. The mice were maintained in ventilated filter cages under specific pathogen-free conditions and fed ad libitum. MMTV-*NCAPH* mice were generated at the Transgenic Facility of the University of Salamanca. A 2226 bp fragment containing the entire human *NCAPH* coding region was generated by PCR, cloned, and inserted into

the *EcoRI* site of the pMKbpAII plasmid containing the Mouse Mammary Tumour Virus (MMTV) promoter. The construct containing *NCAPH* free of any vector sequence (BssHII fragment) was injected into fertilised oocytes extracted from FVB animals. Mice were screened for the presence of the transgene in the Southern blots of tail DNA digested with *XhoI*. The blots were hybridised with the same *XhoI* DNA fragment (658 bp) and confirmed by qPCR performed on mammary gland tissue of 3-month-old mice. Two founders (*NCAPH #1* and *NCAPH #2*) were obtained and bred, and their tail DNAs were genotyped using PCR. Three MMTV-*NCAPH* cohorts were generated: one cohort of nulliparous mice ($N = 10$) and two cohorts of parous mice that were pregnant twice: MMTV-*NCAPH* #1 ($N = 29$) and *NCAPH* #2 ($N = 29$). FVB/N-Tg(MMTVneu)202Mul/J mice carrying the *avian erythroblastosis oncogene B2/neuroblastoma-derived* (*ErbB2/cNeu*) protooncogene (*ErbB2* after that) under the control of the MMTV promoter (MMTV-*Erbb2* transgene)[34] were obtained from Jackson Laboratory. A new cohort of mice was obtained by crossing MMTV-*NCAPH* and MMTV-*ErbB2* mice to obtain dual transgenic mice ($N = 33$).

2.3 | Backcrossing and F1 allografts

A genetically heterogeneous mouse cohort was generated by backcrossing two inbred strains as previously described.[35] Briefly, we crossed the breast cancer-resistant C57BL/6 mouse strain (C57) with an *FVB/N-Tg(MMTVneu)202Mul/J* susceptible strain (FVB). F1-*Neu*$^+$ males generated with the transgene were mated with FVB non-transgenic females to obtain a backcrossed cohort of MMTV-*ErbB2* mice (BX-*Neu*$^+$) ($N = 147$). The concentration of tail DNA was measured using a Nanodrop ND-1000 Spectrophotometer and used for genotyping.[35]

Tumour cells from BX-*Neu*$^+$ mice were transplanted into F1 female recipients and 100 μL of a single-cell suspension containing $2-5 \times 10^6$ cells was injected into both inguinal flanks of each mouse. Each tumour was transplanted into two individuals ($N = 125$). Chemotherapy was initiated when the tumour diameter reached 12 mm by treating 58 mice with docetaxel (25 mg/kg; Taxotere, Sanofi Aventis) and 69 mice with doxorubicin (5 mg/kg; Farmiblastina, Pfizer), and each drug was injected intraperitoneally (IP). Docetaxel and doxorubicin were administered every 8 and 10 days, respectively, and the mice were sacrificed when the tumour reached 25 mm in diameter or 2 months after the end of treatment.

2.4 | Histological analysis

Tumours and mammary glands were fixed in 4% paraformaldehyde (PFA: Scharlau FO) for 24 h at room temperature and washed in 70% ethanol before being embedded in paraffin for automated processing (Shandon Excelsior, Thermo). The samples were sectioned and stained with haematoxylin and eosin to evaluate their pathology under a microscope. Five photos (10× magnification) were taken randomly with a Leica ICC50 HD camera under the control of Leica Application Suite V3.7 software, quantifying the relative ductal area of the mammary glands. Image analysis was performed using *ImageJ*, selecting the ductal area (epithelial cells forming the duct) while excluding the adipose tissue and duct lumen. The ductal epithelial area was divided by the total field area to calculate the relative percentages, and the mitotic index of the tumours was defined as very high if there were more than eight mitoses at 40× magnification, high if there were between four and eight mitoses, moderate when there were between two and four mitoses, and low if there were less than two mitoses. A pathologist evaluated this parameter in the Pathology Unit of our Centre.

2.5 | Immunostaining of mouse tissue

Immunostaining of mouse tissues was performed at the Pathology Unit of our Centre. Mammary gland or tumour sections (3 μm) were deparaffinised and probed with a primary antibody against Ki-67 (MAD020310Q at a 1:50 dilution: Master Diagnostica) using Discovery ULTRA (Roche). The secondary antibody used was OmniMap anti-Rb horseradish peroxidase (HRP) (#05269679001, Roche), and Ki-67$^+$ cells were quantified using Leica Application Suite V3.7 software with five selected areas on the slide at 20× magnification.

2.6 | Cell culture

MCF-7 and BT549 cells were grown in complete DMEM containing 4.5 g/L glucose and L-glutamine (Sigma–Aldrich) at 37°C in a humid atmosphere containing 5% CO_2. The medium was supplemented with 56 IU/mL penicillin, 56 mg/L streptomycin (Invitrogen), and 10% foetal bovine serum (LINUS #16sV30180.03), and all cell cultures were routinely tested for mycoplasma contamination. The cell lines were analysed for authentication at the Genomics Core Facility at the Instituto de Investigaciones Biomédicas 'Alberto Sols' (CSIC-UAM) using STR PROFILE DATA, the STR amplification kit (GenePrint 10 System, Promega),

STR profile analysis software GeneMapper v3.7 (Life Technologies), and a Genomic Analyser System ABI 3130 XL (Applied Biosystems). See the Supporting Information section for procedures related to cell viability, soft agar assays, Boyden chamber cell migration assays and qPCR assays.

2.7 | Protein analysis

Proteins were analysed by Western blotting as previously described.[35] In brief, proteins were extracted from the tumours in RIPA buffer (150 mM NaCl, 1% [v/v] NP40, 50 mM Tris–HCl [pH 8.0], .1% [v/v] SDS, 1 mM EDTA, .5% [w/v] deoxycholate), whereas the proteins from the cell lines were extracted in TNES buffer (100 mM NaCl, 1% [v/v] NP40, 50 mM Tris–HCl [pH 7.6], 20 mM EDTA), both buffers containing protease and phosphatase inhibitor cocktails (Sigma–Aldrich; #P8340). The recovered proteins were quantified using a Bradford Protein Assay (Bio-Rad, #5000006), resolved by SDS-PAGE on 10 or 12% gels (Bio-Rad, #456-8085), and transferred to polyvinylidene difluoride membranes (Immobilon-P, Millipore). The membranes were probed with the following primary antibodies raised against NCAPH (1:10000, #TA303239: OriGene), cyclin D1 (1:1000, #sc8396: Santa Cruz Biotechnology), γH2AX (1:200, Ser139 #05-636: Millipore), pCHK1 (1:200, Ser345 #2348: Cell Signalling), tubulin (1:1000, DM1A; Sigma–Aldrich; #T6199), HSP90 (1:1000, #515081: Santa Cruz Biotechnology), actin (1:1000, #A5441: Sigma–Aldrich), pERK1/2 (1:1000, #9101: Cell Signalling), total ERK1/2 (1:1000, #4696: Cell Signalling), pAKT S473 (1:1000, #32581: Elabscience), total AKT (1:1000, #30471: Elabscience), cleaved Caspase-3 (Cell Signalling #9661), anti-ERBB2 (1:1000, #ab2428: Abcam) and anti-phospho-ERBB2 (Tyr1248) (1:1000, #ab47755: Abcam). Antibody binding was detected using HRP-conjugated anti-mouse or anti-rabbit secondary antibodies (1:10,000 dilution; Amersham) and visualised by enhanced chemiluminescence (ECL, #170-5061: Bio-Rad). Images were acquired using an ImageQuant LAS 500 Chemiluminescence CCD camera (GE Healthcare Life Sciences).

2.8 | Inducible system for NCAPH

An inducible mammalian expression construct encoding NCAPH was obtained by cloning *NCAPH* into a pRetroX-Tight-Puro vector (Clontech #632104). The MCF-7 and BT-549 inducible *NCAPH* systems were generated by co-transfecting pRetroXTet-On Advanced with pRetroX-Tight-Puro-*NCAPH*. *NCAPH* expression was induced in cells after exposure to doxycycline (10 μg/mL).

2.9 | Identification of differentially expressed genes and functional enrichment analysis

The quality and quantity of the total RNA isolated from the cells were determined using an Agilent 2100 Bioanalyser and NanoDrop ND-1000. Affymetrix GeneChip mouse gene 1.0 ST arrays were used, according to the manufacturer's protocol. Gene expression data for mouse breast cancers are available from the Gene Expression Omnibus (accession number GSE54582).[35] The differentially expressed genes (DEGs) between animals with *Ncaph* high and *Ncaph* low in the BX-*Neu*+ cohort were identified using Transcriptome Analysis Console software, using a cutoff change (|log FC| > 2.0) and adjusted $p \leq .05$. Gene Ontology (GO) pathway enrichment analyses were performed using the R package 'clusterProfiler'.[36,37]

The DEGs obtained in BX-*Neu*+ mice were analysed by LASSO regression using the R package 'glmnet',[31] and the lambda value was determined by cross-validation to penalise collinearity among genes. The animals were divided into three groups (high, medium and low risk) and their scores were calculated using the following equation:

$$Risk\ score = \sum_{i=0}^{n} exp_i \times \beta_i$$

where exp_i indicates each genes's expression level. Kaplan–Meier (KM) lifespan analysis was performed to compare the prognostic differences between the three groups (R package: 'survival'[38,39] and 'survminer'[40]). For human analysis, we studied a combined patient database using Gene Expression-based Outcome for Breast Cancer Online (GOBO),[41] an online tool that downloads gene expression levels from an 1881-sample breast cancer dataset. We used the GSE1456, GSE2603, GSE6532, GSE3494, GSE4922, GSE6532, GSE7390, GSE11121, GSE12093, GSE2034 and GSE5327 databases to select luminal A samples (401 patients). We first conducted a univariate Cox regression analysis of DEGs and then selected candidate genes related to relapse-free survival (RFS) according to a criterion of *p*-value <.25.[42,43] In multivariate analysis, a lenient *p*-value threshold (e.g., $p \leq$ .25) is often used for variable selection, as advocated by Dr. Frank E. Harrell Jr. in 'Regression Modeling Strategies'.[43] This approach ensures that potentially significant predictors are not excluded prematurely despite their weaker associations in bivariate analyses. This inclusive initial selection, followed by refinement through techniques such as stepwise regression, aims to create a model that is comprehensive yet parsimonious, effectively capturing data relationships. The selected genes obtained in the univariate analysis were further analysed by LASSO

regression using glmnet, and the regression coefficient lambda was determined by cross-validation. Specifically, the GOBO cohort was randomly split into a training set (70%) to develop the model and a test set (30%) for validation. We then identified 10 genes that helped define disease prognosis using the LASSO model, which we referred to as the Gene Signature for Luminal A 10 (GSLA10). For the KM analysis of RFS, the patients were divided into three groups (high, medium and low risk), and the score was calculated using the aforementioned equation to compare the prognostic difference between the three groups. The results were then verified using receiver operating characteristic (ROC) curve analysis.[44] Finally, we validated our model using two independent databases, Molecular Taxonomy of Breast Cancer International Consortium (METABRIC, 718 luminal A breast cancer patients)[45] and The Cancer Genome Atlas Breast Invasive Carcinoma (TCGA-BRCA, 499 luminal A breast cancer patients),[46] both of which are available from cBioPortal (https://www.cbioportal.org/). In comparison with the Oncotype DX signature, we also constructed an Oncotype DX model. According to Paik et al., the expression levels of each Oncotype gene were normalised against a set of five reference genes: *ACTB*, *GAPDH*, *GUS*, *RPLPO* and *TFRC*. This normalisation involved measuring each gene's expression and then adjusting it by subtracting the mean expression value of these five reference genes.[6] In both the training cohort (GOBO) and independent validation cohorts (METABRIC and TCGA-BRCA), we evaluated, validated and compared the ability of GSLA10 and Oncotype to define RFS (GOBO) and DFS (METABRIC and TCGA-BRCA) in luminal A tumours. During independent validation, both the GSLA10 and Oncotype pre-trained models from GOBO were deployed without any modifications.

The concordance index (*C*-index) is a statistical tool used to assess the predictive accuracy of a prognostic model, particularly in survival analysis. It is calculated by forming all possible pairs of subjects, evaluating whether the predicted survival times align with the actual survival outcomes, and excluding pairs with indeterminate ordering due to censoring or ties. The *C*-index is the ratio of concordant pairs to the total number of evaluable pairs. Values range from .5 (no predictive ability) to 1.0 (perfect prediction), with higher values indicating better model performance.[47,48]

2.10 | Statistical analysis

Depending on the data distribution, we calculated either the Pearson or Spearman correlation coefficients or performed a Student's *t*-test or Mann–Whitney *U*-test to compare continuous variables between the two groups.

ANOVA, or the Kruskal–Wallis test, was used to compare continuous variables across more than two groups. We used the KM estimator and log-rank test to compare temporal variables. For contingency analysis, we used Fisher's exact test to analyse 2 × 2 tables and the chi-square test in other cases.

For more material and methods data, see the Supporting Information section.

3 | RESULTS

3.1 | NCAPH overexpression induces mammary gland hyperplasia and breast cancer in mice

NCAPH is a constituent of the condensin complex, which facilitates the assembly of interphase chromatin into chromosomes and orchestrates their segregation during mitosis and meiosis[21] (Figure 1A). Initially, we found that NCAPH overexpression was associated with poor outcomes in breast cancer[49,50] (personal communications). This was determined through data mining analyses, specifically examining genes involved in mitosis and their correlation with adverse prognosis in breast cancer. We sourced our data from multiple databases: ETAM-158,[51] GSE1456,[52] GSE2034,[53] GSE4922[54] and the dataset cited in Ref.[55] (Figure S1A). Later, we identified an elevation in NCAPH levels in invasive ductal breast carcinomas compared to normal mammary tissues[56] (Figure 1B). Consequently, this prompted us to investigate the potential involvement of NCAPH in the pathogenesis of breast cancer.

To determine whether NCAPH is a driver of breast cancer development, we generated transgenic mice overexpressing NCAPH under the control of the MMTV promoter (Figure 1C), inducing NCAPH overexpression in the mammary gland (Figure 1D,E). MMTV-*NCAPH* nulliparous mice developed breast tumours during long-term follow-up, with 30% of mice developing breast cancer after 120 weeks (Figure 1F). Hence, the overexpression of *NCAPH* could drive breast cancer development. Oncogenes driven by the MMTV promoter are overexpressed during pregnancy because of promoter induction by gravidity hormones, resulting in more aggressive tumours.[57] Indeed, the overexpression of NCAPH, regulated by the MMTV promoter in this transgenic line, resulted in breast tumour development with reduced latency postpregnancy, suggesting a dose-dependent effect (Figure 1G). Both *MMTV-NCAPH#1* and #2 mouse transgenic lines developed breast tumours after pregnancy, with no significant difference in tumour incidence (Figure 1H). Infiltrating ductal carcinomas generated by *NCAPH* overexpression had a range of histopathological features,
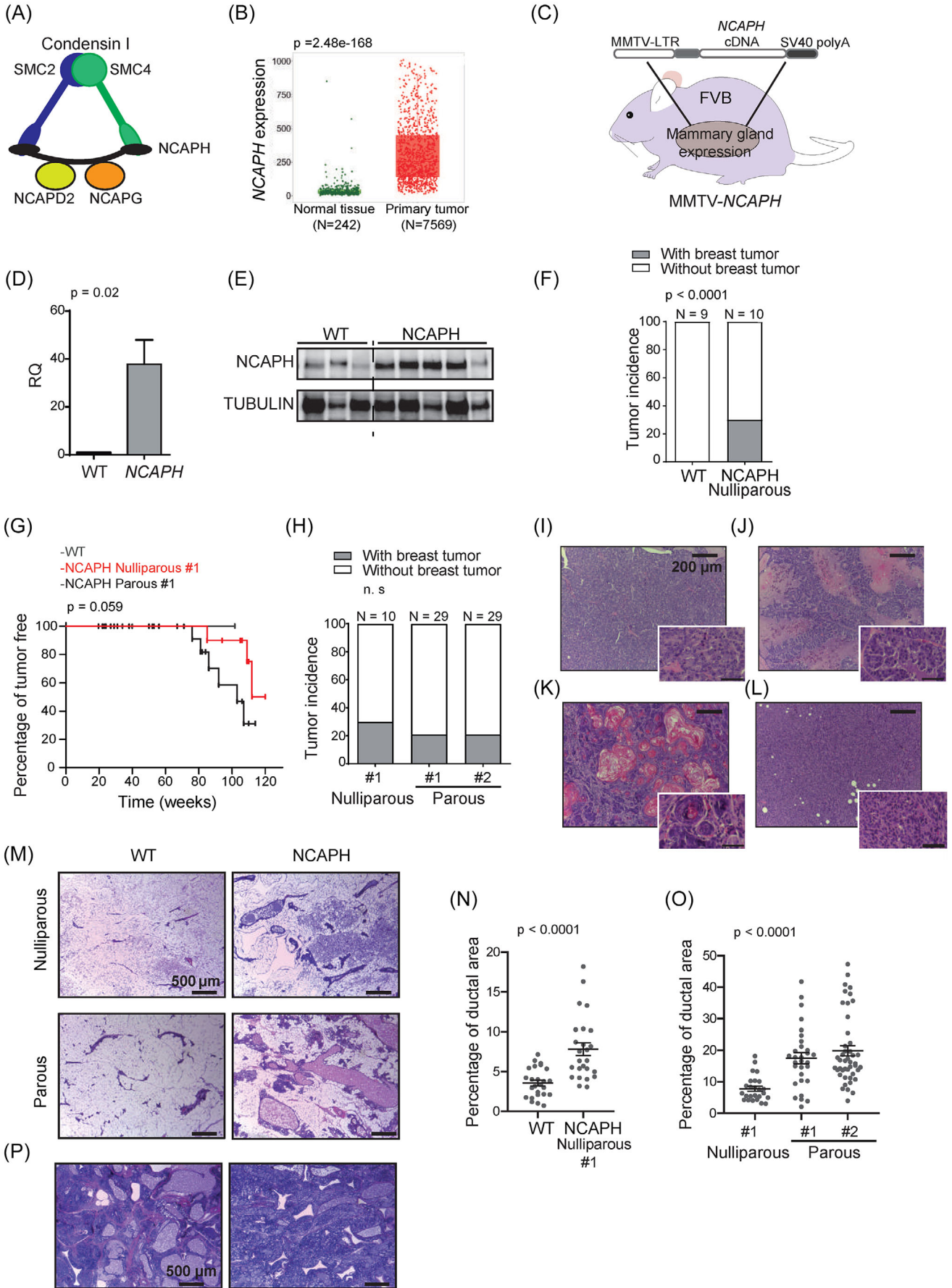
including breast adenocarcinomas with papillary differentiation, squamous differentiation, or a mesenchymal pattern (Figure 1I–L and Table S1).

Interestingly, although most nulliparous *MMTV-NCAPH* mice did not develop breast tumours after 2 years, they did show a significant increase in their ductal epithelial components. Indeed, *MMTV* promoter-driven overexpression of *NCAPH* produced hypertrophic mammary glands with marked ductal hyperplasia. The ducts were formed by a normal, single row of epithelial cells with no atypia and a benign aspect (Figure 1M). However, there was a substantial increase in the number of ducts, increasing the total parietal ductal area (Figure 1N,O). Notably, two additional mice that were not included in the comparison exhibited massive ductal and stromal hyperplasia in the mammary gland and the absence of fatty tissue (Figure 1P). Together, these results indicate that NCAPH is oncogenic in breast tumours.

3.2 | NCAPH expression is associated with poor outcome and response to therapy in luminal A breast cancer patients

After demonstrating that NCAPH is involved in breast cancer development (Figure 1) and is linked to poor prognosis in breast cancer (Figure S1A), we aimed to identify the specific intrinsic subtype of breast cancer where intratumoural NCAPH levels are associated with unfavourable outcomes. Interestingly, the most aggressive subtypes of breast cancer—specifically basal and HER2-enriched—exhibited the highest levels of NCAPH (Figure 2A).[58] However, elevated NCAPH expression was also observed in patients from other subgroups. Consequently, we explored if NCAPH levels could differentiate between prognostically favourable and unfavourable forms within each intrinsic breast cancer subtype.

Breast cancer subtype classification and prognosis have been enhanced by gene signatures such as Prediction Analysis of Microarray 50 (PAM50).[9] PAM50 provides a more precise intrinsic subtype classification than the St. Gallen approach, especially evident in the distinction between luminal A and B subtypes, underscoring its superior molecular accuracy.[60] Consequently, we employed PAM50 to explore whether distinct evolutionary groups emerge from intratumoural NCAPH expression under optimal classification conditions.[59]

Although *NCAPH* expression was weak in the luminal A subtype compared with some other subtypes (Figure 2A), we found that patients with high levels of *NCAPH* RNA were associated with poor outcome ($p = 4.8 \times 10^{-7}$). Paradoxically, basal-like tumours were associated with a good clinical outcome ($p = .0066$) (Figure 2B).

These results indicate the existence of a subpopulation of luminal A tumours that have poor outcome, which can be distinguished by their high intratumoural levels of *NCAPH*. We evaluated NCAPH levels by IHC in a cohort of patients with luminal A tumours (Table S2A), and again, we confirmed the higher NCAPH protein levels in patients with a poor outcome (presence of liver metastases) than in those who evolved well in the 10-year follow-up. The median RI for the good prognostic group was 25.06%, with an interquartile range (IQR) of [18.94–30.60]. For the poor prognostic group, the median RI was 43.06%, accompanied by an IQR of [24.74–56.44]. Across the entire cohort, the median RI was observed to be 28.73%, with an IQR of [19.45–34.40] (Figure 2C,D). Moreover, NCAPH levels were not associated with other tumour characteristics such as grade, stage, histological subtype, or Ki-67 staining (Table S2B). Significantly, high levels of *NCAPH* in luminal A tumours were associated with poor response to endocrine therapy and chemotherapy in the long-term follow-up (Figure 2E). Thus, elevated *NCAPH* levels in luminal A tumours from patients who have received hormonal therapy or chemotherapy correlate with poor long-term outcomes, thus serving as a prognostic marker. This was in contrast with other intrinsic tumour subtypes that responded to therapy independently of *NCAPH* levels

FIGURE 1 Transgenic mice that overexpress non-SMC condensin I complex subunit H (NCAPH) develop breast cancer and mammary gland hyperplasias. (A) Condensin complex 1, of which NCAPH forms a part. (B) *NCAPH* is more strongly expressed in breast cancer than in normal mammary tissue. Analysis performed in the UALCAN portal using The Cancer Genome Atlas (TCGA) information. TPM, transcripts per million. (C) The figure shows *NCAPH* overexpressed under the Mouse Mammary Tumor Virus (MMTV) promoter to generate MMTV-*NCAPH* transgenic mice. (D and E) Overexpression of *NCAPH* RNA in the mammary gland of transgenic mice quantified by qPCR (D) and the increase in protein seen in Western blots (E). (F) Incidence of breast cancer in mice overexpressing NCAPH after a 120-week follow-up. The controls were non-transgenic mice with an FVB genetic background. (G) Tumour latency of MMTV-*NCAPH*#1 nulliparous and parous mice. Kaplan–Meier curves and log-rank test. (H) Incidence of breast cancer in parous mice overexpressing NCAPH after a 120-week follow-up. (I–L) Different histopathological patterns of breast cancer in MMTV-*NCAPH* transgenic mice (Table S1). Glandular pattern (I), papillary pattern (J), squamous pattern with corneal pearls (K), solid pattern (L). (M) Ductal hyperplasia in nulliparous and multiparous MMTV-*NCAPH* mice, with ductal dilation. (N and O) Increased ductal parietal area of the mammary gland in MMTV-*NCAPH* nulliparous versus wild-type mice (N). The multiparous mice had a larger ductal parietal area than the nulliparous mice (O). N and O, *t*-test. (P) Detail of a mammary gland from two MMTV-*NCAPH* mice with massive hyperplasia. SMC, structural chromosome maintenance.
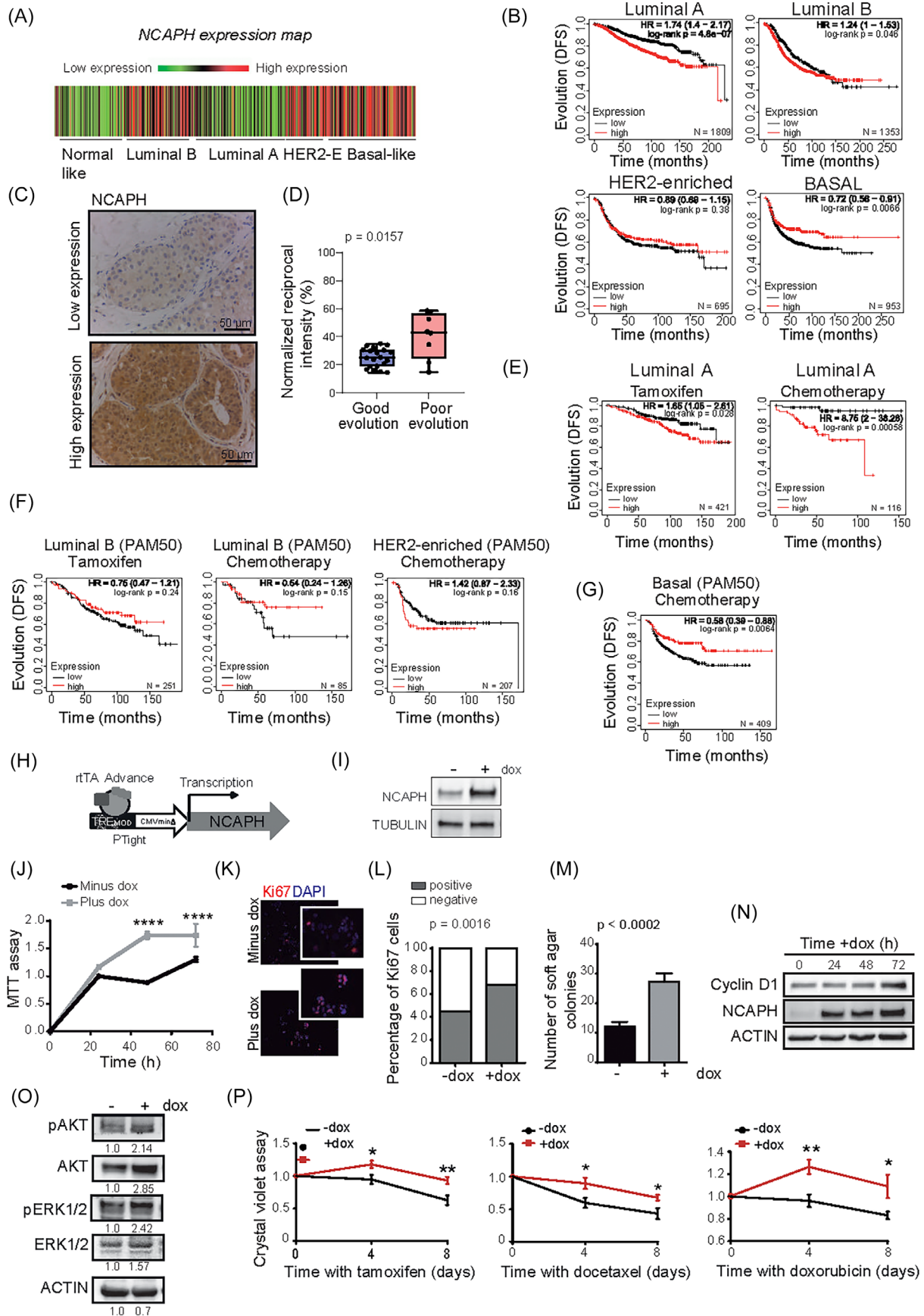
FIGURE 2   Intratumoural levels of non-SMC condensin I complex subunit H (NCAPH) and outcome of breast cancer. (A) Representation of *NCAPH* expression in the various intrinsic subtypes of breast cancer. Data were obtained from the bc-GenExMiner.[58] (B)

(Figure 2F), except for basal tumours, which tended to respond well to chemotherapy when *NCAPH* levels were high (Figure 2G).

To ascertain that elevated levels of NCAPH were associated with a diminished response to chemotherapy, we engineered an MCF7 luminal A breast cancer cell line wherein NCAPH overexpression could be triggered by doxycycline (Figure 2H,I). The induction of NCAPH overexpression in MCF7 cells augmented their viability and proliferation (Figure 2J–L), which correlated with the generation of significantly more colonies on soft agar (Figure 2M) and elevated levels of cyclin D1 expression (Figure 2N). The elevated levels of NCAPH were correlated with an increase in both total and phosphorylated AKT levels, which play a pivotal role in regulating various cellular processes, including proliferation, survival and cell growth (Figure 2O). Remarkably, the upregulation of NCAPH led to partial resistance to therapy in MCF7 cells (Figure 2P). We observed no discernible differences in cell viability in the basal BT-549 cell line following the induction of NCAPH expression or doxorubicin treatment (Figure S2A–C). Thus, high NCAPH expression in luminal A breast tumours is associated with a poorer prognosis and therapy response, indicating its potential as an identifier for high-risk luminal A tumours.

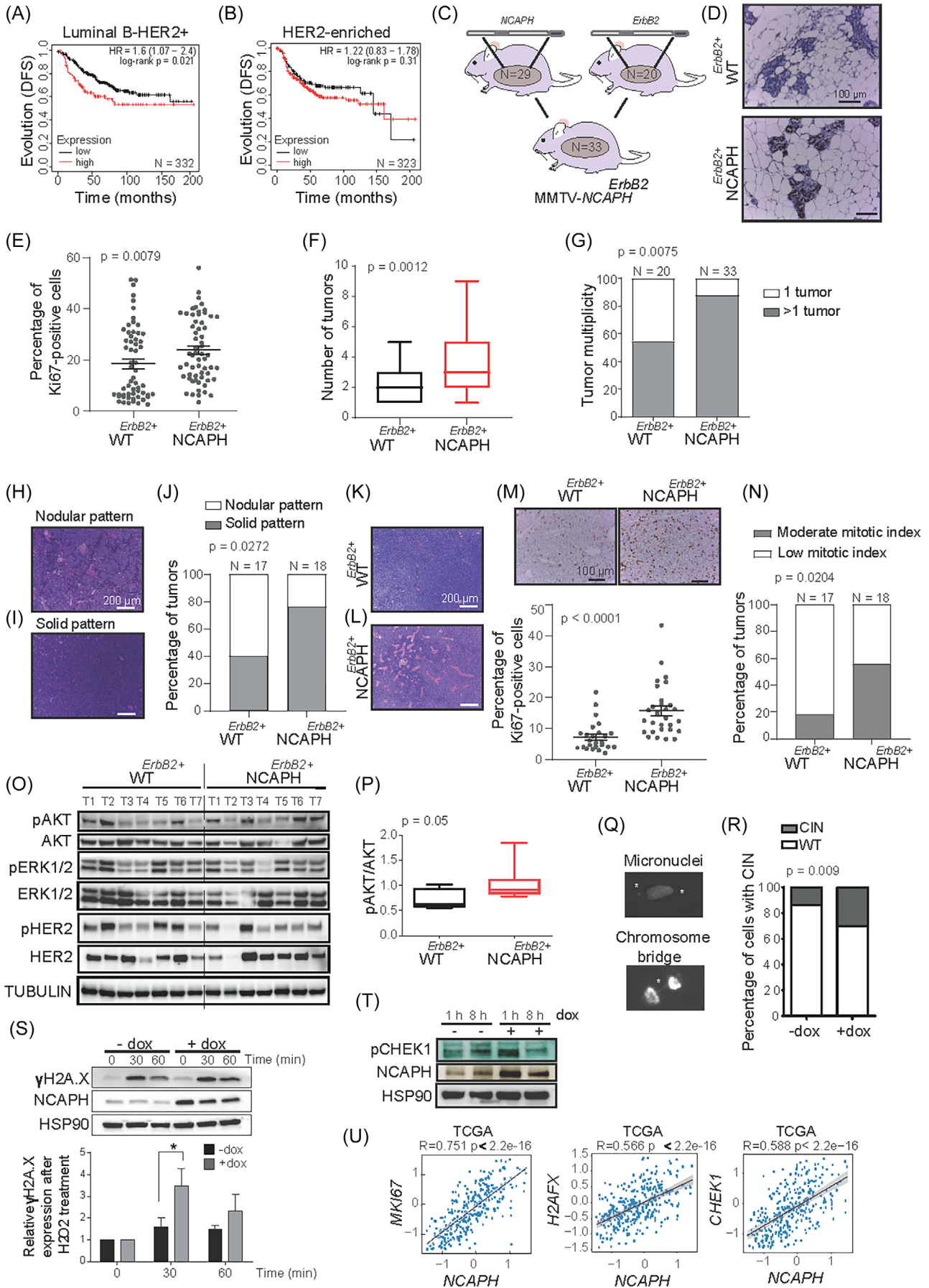## 3.3 | Intratumoural levels of NCAPH are associated with poor outcome in luminal HER2[+] tumours

NCAPH levels were not associated with changes in the outcome of HER2-enriched tumours, which are ER negative, as defined by the PAM50 (Figure 2A,E). Since luminal HER2 tumours are not defined as an intrinsic subtype of breast cancer defined by PAM50, immunohistochemical classification was used to define the role of NCAPH in the prognosis of this tumour subtype.[61] Interestingly, high *NCAPH* levels were associated with poor clinical outcome of HER2[+] luminal tumours (ER positive), confirming that *NCAPH* did not influence the outcome of HER2-enriched tumours defined by IHC either (Figure 3A,B).

Thus, to investigate the role of NCAPH in the pathogenesis of luminal-HER2[+] tumours, we crossed MMTV-*NCAPH* mice with MMTV-*ErbB2* transgenic mice that developed luminal ERBB2 breast tumours[34] (Figure 3C). An increase in epithelial proliferation was evident in the mammary glands of double-transgenic MMTV-*NCAPH*[ErBb2] mice relative to their MMTV-*ErbB2* counterparts when Ki-67[+] cells were stained by IHC and counted (Figure 3D,E). Interestingly, the overexpression of *NCAPH*[ErBb2] in the mammary tissue of mice (MMTV-*NCAPH*[ErBb2]) developed significantly more tumours than their MMTV-*ErbB2* counterparts (Figure 3F,G), although no differences were observed in any other pathophenotypes of this disease.

Notably, while *NCAPH* transgenic mice developed distinct histopathological types of breast cancer (Figure 1J–M), the MMTV-*NCAPH*[ErBb2] double-transgenic mice developed only infiltrating ductal adenocarcinomas, suggesting that the *ErbB2* oncogene exerts a dominant effect on the tumour phenotype. Therefore, the *ErbB2* oncogene appears to reprogram tumour differentiation so that instead of the histopathologically distinct tumours induced by NCAPH overexpression, the characteristics of infiltrating ductal adenocarcinoma predominated.

Outcome (disease-free survival [DFS] probability) of patients with different intrinsic subtypes of breast cancer, as defined by Prediction Analysis of Microarray 50 (PAM50). The selection of patients for this panel does not consider the type of treatment received. (C) Analysis of NCAPH expression by immunohistochemistry (IHC) in luminal A tumours. The panel shows an example of strong and weak expressions. (D) Comparative analysis of NCAPH staining intensity in luminal tumours exhibiting good versus poor prognosis was performed using Fiji Is Just ImageJ (FIJI) software, adhering to Nguyen et al. methodology. Post-colour deconvolution for DAB staining, reciprocal intensity (RI) in arbitrary units was determined by the difference in mean intensity between the epithelium and the background. These RI values, which inversely correlate with staining intensity, were normalised to the background intensity to ensure precise representation. (E) Outcome (DFS) in luminal A patients (as defined by PAM50) treated with tamoxifen or chemotherapy. (F) Patients with luminal B or HER2-enriched tumours were treated with tamoxifen and chemotherapy. (G) Basal tumours (as defined by PAM50) were treated with chemotherapy. In (B), (F) and (G), the panels were generated with the Kaplan–Meier Plotter database, considering high and low intratumoural levels of *NCAPH* mRNA defined by the best cutoff.[59] (H) Diagram of the TET on system used to express NCAPH in response to doxycycline (dox). (I) Expression of NCAPH in the MCF-7 cell line after a 24 h induction with dox. (J) Assessment of cell proliferation in the MCF-7 cell line following NCAPH induction using the MTT test. (K) Representative image of Ki-67 staining (red) and DAPI (blue) after a 24 h induction with dox in the inducible MCF-7 NCAPH cell line. (L) Quantitative analysis of proliferative cells positive for Ki67 using Fisher's exact test. (M) NCAPH induction led to an increase in the number of MCF-7 colonies formed in soft agar. (N) A Western blot illustrates the upregulation of cyclin D1 post-NCAPH induction. (O) Analysis of AKT and ERK signalling pathways in the MCF-7 cell line post-NCAPH induction as evidenced by Western blotting. Densitometric quantification of each protein is displayed below its respective band. (P) Increased cell resistance to various treatments, including tamoxifen, docetaxel and doxorubicin, observed after NCAPH induction. The number of patients included in each Kaplan–Meier curve panel is indicated within each respective graph. SMC, structural chromosome maintenance.

(A) Luminal B-HER2+
HR = 1.6 (1.07 – 2.4)
log-rank p = 0.021
N = 332

(B) HER2-enriched
HR = 1.22 (0.83 – 1.78)
log-rank p = 0.31
N = 323

(C) *NCAPH* N=29 *ErbB2* N=20 *ErbB2* MMTV-*NCAPH* N=33

(D) *ErbB2+* WT *ErbB2+* NCAPH 100 µm

(E) p = 0.0079

(F) p = 0.0012

(G) p = 0.0075 N = 20 N = 33 1 tumor >1 tumor

(H) Nodular pattern 200 µm

(I) Solid pattern

(J) Nodular pattern Solid pattern p = 0.0272 N = 17 N = 18

(K) *ErbB2+* WT 200 µm

(L) *ErbB2+* NCAPH

(M) *ErbB2+* WT NCAPH 100 µm p < 0.0001

(N) Moderate mitotic index Low mitotic index p = 0.0204 N = 17 N = 18

(O) *ErbB2+* WT *ErbB2+* NCAPH T1 T2 T3 T4 T5 T6 T7 T1 T2 T3 T4 T5 T6 T7 pAKT AKT pERK1/2 ERK1/2 pHER2 HER2 TUBULIN

(P) p = 0.05

(Q) Micronuclei Chromosome bridge

(R) CIN WT p = 0.009

(S) - dox + dox Time (min) γH2A.X NCAPH HSP90 -dox +dox

(T) 1 h 8 h 1 h 8 h dox pCHEK1 NCAPH HSP90

(U) TCGA R=0.751 p< 2.2e−16 MKI67 NCAPH
TCGA R=0.566 p < 2.2e−16 H2AFX NCAPH
TCGA R=0.588 p< 2.2e−16 CHEK1 NCAPH

Moreover, the double-transgenic mice predominantly developed infiltrating ductal adenocarcinomas exhibiting a solid histopathological pattern (Figure 3H–J), enhanced vascularisation, and significantly heightened tumour proliferation compared to their MMTV-*ErbB2* counterparts (Figure 3K–M). Remarkably, tumours induced by NCAPH overexpression demonstrated a higher mitotic index (Figure 3N).

The different histopathological behaviors of tumours from MMTV-*NCAPH*^*ErbB2* double-transgenic mice prompted us to study molecules from some of the main pathways downstream of ERBB2. Intratumoural signalling was evaluated by assessing pAKT/pmTOR and pERK in Western blots of seven tumours from each phenotype, and only pAKT levels were significantly higher in tumours from the double-transgenic mice than in those from the single-transgenic mice (Figure 3O,P). No significant differences were observed in the pERBB2 and ERBB2 levels between tumours from MMTV-*ErbB2* mice and double-transgenic MMTV- *Ncaph*^*ErbB2* mice (Figures 3O and S3A,B), ruling out their involvement in the differing tumour behaviors of the two groups. Additionally, the tumours from the double-transgenic MMTV-*Ncaph*^*ErbB2* mice exhibited higher basal apoptosis, as indicated by elevated levels of cleaved Caspase 3 (Figure S3C).

It is noteworthy that NCAPH overexpression in MCF7 cells also elevated the proportion of cells exhibiting chromosomal instability (CIN), wherein micronuclei and chromosomal bridges were identifiable (Figure 3Q,R).

Genomic instability triggers genomic stress,[63,64] as evidenced by the escalated levels of $\mathfrak{o}$H2AX and pCHEK1 following NCAPH overexpression and exposure to H_2O_2 (Figure 3S,T). In human tumours, a robust positive correlation was discerned between NCAPH expression and proliferation markers (Ki67) or genomic instability markers (H2AX and CHEK1) (Figure 3U).

The unique histopathological characteristics observed may emanate from elevated levels of NCAPH in the mammary gland, thereby exacerbating genomic instability and instigating secondary oncogenic events that target diverse tumour differentiation pathways.[65]

Together, these findings propose that NCAPH overexpression in ERBB2 tumours engenders more aggressive histopathology, solid tumours with high mitotic indices, and enhanced vascularisation and cell proliferation. Thus, elevated levels of NCAPH promote a more aggressive form of breast cancer, potentially elucidating the adverse progression observed in luminal tumours.

3.4 | NCAPH expression is associated with poor breast cancer outcome in a genetically heterogeneous cohort of mice

Genetically heterogeneous mouse cohorts better reflect the heterogeneity observed in the human population, facilitating the identification of the genetic and transcriptomic determinants associated with disease outcome.[35,66,67]

FIGURE 3 Effect of non-SMC condensin I complex subunit H (NCAPH) expression on luminal ERBB2-positive tumours in vivo. (A and B) Overexpression of *NCAPH* in patients with luminal ER-positive HER2+ breast tumours was associated with a worse outcome (A), while this association was not seen in HER2+ ER-negative tumours (B). Panels A and B were generated in the KM Plotter database.[59] Each graph displaying a Kaplan-Meier curve panel specifies the number of patients it encompasses. Outcome (disease-free survival [DFS] probability). Analysis using the KM plotter's optimal cutoff. (C) Scheme to generate double-transgenic mice Mouse Mammary Tumor Virus (MMTV)-*NCAPH*^*ErBb2*. (D and E) Immunohistochemical proliferation assessment in non-tumoural mammary glands: analysis of mammary glands from MMTV-*ErbB2* (upper panels) and MMTV-*NCAPH*^*ErbB2* double-transgenic mice (D), with representative Ki67 immunohistochemistry images. Five mice per group were analysed, and field analysis results are shown in panel E, using the Mann–Whitney *U*-test. (F and G) Tumour development quantification: double-transgenic mice exhibited significantly more tumours (F) and higher tumour multiplicity (G) compared to MMTV-*ErbB2* mice. (H–J) Histopathological pattern analysis: this section assesses different histopathological patterns in the samples, focusing on nodular (H) and solid patterns (I). Notably, the solid pattern occurred more frequently in double-transgenic mice compared to MMTV-*ErbB2* mice (J). Ten mice per group were evaluated with a chi-squared test. (K and L) Vascularisation variability in mouse tumours: the panels compare vascularisation in tumours from MMTV-*ErbB2* and MMTV-*NCAPH*^*ErbB2* double-transgenic mice. Haematoxylin–eosin staining is used to emphasise vascular differences between MMTV-*ErbB2* (panel K) and MMTV-*NCAPH*^*ErbB2* double-transgenic mice (panel L). (M) Proliferative activity in tumour tissues: this section uses Ki67 immunohistochemical analysis to evaluate tumour proliferation in MMTV-*ErbB2* (left) and MMTV-*NCAPH*^*ErbB2* double-transgenic mice (right). Quantitative results from this analysis are displayed below the corresponding images. (N) Mitotic index quantification using chi-squared test. (O and P) Analysis of ERBB2 downstream signalling molecules via Western blot (O) with pAKT quantification (P). (Q and R) NCAPH induction increased genomic instability, evidenced by more micronuclei and chromosomal bridges (CIN). Micronuclei and chromosomal bridge images (Q) are accompanied by CIN quantification (R), analysed with Fisher's exact test. (S) Genomic stress linked to increased NCAPH is shown by elevated $\mathfrak{o}$H2AX levels post-H_2O_2 exposure and during 30- or 60-min recovery. (T) pCHEK1 expression changes following NCAPH induction are displayed via Western blot. (U) RNA expression correlation in luminal breast cancer: this panel shows the correlation between *NCAPH* RNA and *Ki67*, *CHEK1* and *H2AX* in luminal breast cancer samples from The Cancer Genome Atlas (TCGA) database. It illustrates the correlation coefficients and patterns using CANCERTOOL.[62] SMC, structural chromosome maintenance.
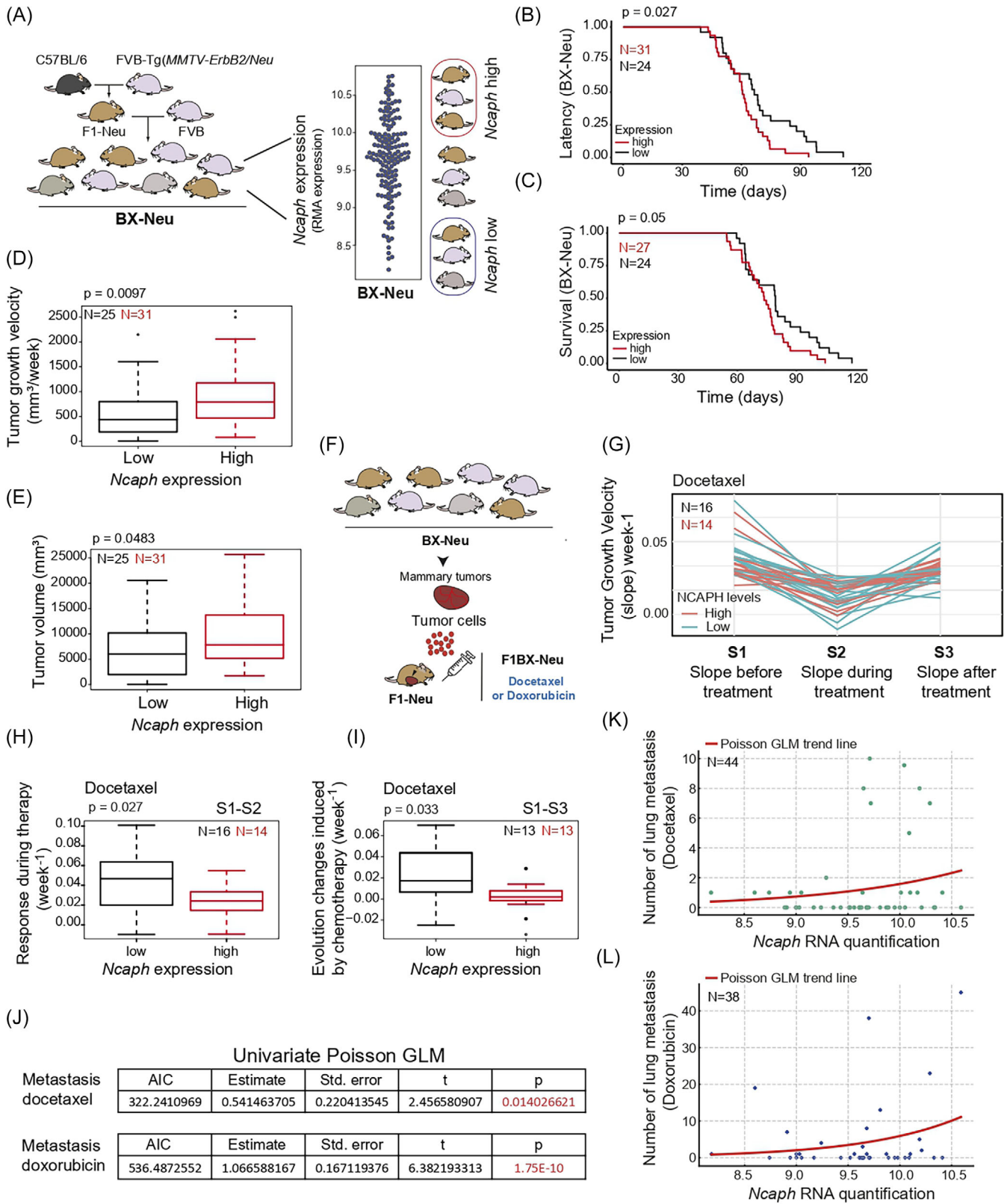
FIGURE 4 Intratumoural levels of *Ncaph* were associated with a worse outcome and response to chemotherapy in a cohort of mice generated by backcrossing (BX-*Neu*⁺). (A) Generation scheme for BX-*Neu*⁺ cohort: the BX-*Neu*⁺ cohort was created by backcrossing transgenic Mouse Mammary Tumor Virus (MMTV)-*ErbB2* mice (FVB/N background) with C57BL/6 mice, which are resistant to breast cancer. The first-generation offspring (F1) were further crossed with wild-type FVB/N mice, producing the BX-*Neu*⁺ cohort. This cohort is genetically diverse, displaying a range of phenotypic traits and intratumoural *Ncaph* gene expression. Mice from the highest and lowest

Thus, to analyse the contribution of NCAPH expression to the heterogeneous outcome of luminal ERBB2 breast cancer, we used a genetically heterogeneous cohort of MMTV-*ErbB2* transgenic mice generated by backcrossing (BX-*Neu*⁺ mice after that) (Figure 4A). We crossed MMTV-*ErbB2* transgenic mice, which are on a susceptible FVB genetic background, with non-transgenic mice on a C57BL/6 genetic background, which is resistant to breast cancer development. We generated F1C57BL6/FVB MMTV-*ErbB2* mice (F1-*Neu*⁺ mice hereafter) that were backcrossed with FVB mice to generate BX-*Neu*⁺ mice, a cohort of mice with more varied breast cancer outcome than genetically homogenous mouse strains.

This BX-*Neu*⁺ cohort was used to evaluate whether intratumoural levels of *Ncaph* were associated with heterogeneous breast cancer outcome, demonstrating that high intratumoural *Ncaph* RNA levels were associated with shorter tumour latency and survival (Figure 4B,C), faster tumour growth, and larger tumour volume (Figure 4D,E). Thus, high levels of *Ncaph* are associated with poor outcome of luminal *ErbB2* breast cancer in the cohort of BX-*Neu*⁺ mice.

Since elevated *NCAPH* levels mediate a poor response to chemotherapy in breast cancer patients and cells (Figure 1G,L–N), we evaluated the response of tumours generated in BX-*Neu*⁺ mice to chemotherapy. Following the laws of transplantation in mice,[68] breast cancer tumours that developed in the backcross cohort were transplanted into F1 mice, and their responses to anthracycline and taxane chemotherapy were evaluated (Figure 4F). This strategy allowed us to evaluate the response of breast cancer to chemotherapy in an extracellular context as homogeneously as possible. Thus, differences in treatment responses can be primarily attributed to differences at the cell-autonomous level. Tumours with high levels of *Ncaph*

responded worse to docetaxel treatment, as reflected by a smaller reduction in tumour size (Figure 4G,H). In addition, the growth rate of tumours with high levels of *Ncaph* was faster, and their outcome was worse after chemotherapy than those that expressed *Ncaph* more weakly (Figure 4G,I). However, *Ncaph* levels did not appear to influence the local response to doxorubicin (Figure S4A–C). After chemotherapy with either doxorubicin or docetaxel, lung metastases were most evident in mice with high intratumoural *Ncaph* levels (Figures 4J–L and S4D,E).

These findings suggest that high intratumoural levels of NCAPH are associated with resistance to breast cancer chemotherapy.

3.5 | A gene signature based on NCAPH expression defines poor tumour outcome in mice and humans

We examined the transcriptomic context in which *Ncaph* expression was associated with poor breast cancer outcome. The wide range of breast cancer outcome in backcross mice[35,63] and the diverse expected patterns of gene expression[67] (Figure 4) make this cohort an excellent tool for identifying transcripts associated with high levels of *Ncaph* expression and poor breast cancer outcome.[35,69] We identified 64 transcripts associated with high intratumoural *Ncaph* levels in breast tumours from the backcross cohort, of which 45 were shared with humans (Figures 5A and S5A,B and Table S3). The functions of this 45-gene signature were assessed using GO enrichment analysis.

Unsurprisingly, several genes associated with high *Ncaph* levels were involved in processes related to correct condensation and segregation of chromosomes during

---

tertiles of *Ncaph* expression were selected for evaluation, as shown in the schema. Gene expression was measured using robust multiarray analysis (RMA) log2. (B and C) Decrease in tumour latency probability (B) and survival probability (C) in BX-*Neu*⁺ mice with elevated intratumoural levels of *Ncaph*. (D and E) Tumour growth velocity and final volume for tumours with high or low levels of *Ncaph*. Weekly tumour volume was measured in mm³ using a digital caliper and calculated as (greater diameter × lesser diameter²)/2. Growth velocity was determined by (final volume – initial volume)/illness duration in weeks, expressed in mm³/week. (F) Diagram showing the allogeneic transplantation strategy, passing tumours generated in BX-*Neu*⁺ mice into immunocompetent F1 mice. Assessments were conducted on mice selected from the top and bottom tertiles. (G–I) Docetaxel Impact on Tumour Growth Relative to Ncaph Levels. (G) This panel displays growth curve slopes at crucial docetaxel treatment stages: pre-chemotherapy (S1) and during chemotherapy (S2, indicating peak response). Response during treatment (RDT) is derived as S1 minus S2, with a higher RDT signifying better response. (H) RDT comparison across tumours with different Ncaph levels, classified by external tertiles. (I) Post-chemotherapy growth changes or Evolution Changes Induced by Chemotherapy (ECIC) are shown by comparing slopes pre-chemotherapy (S1) and post-chemotherapy (S3), with a greater ECIC value indicating improved treatment outcome (external tertiles). (J) Univariate Poisson regression analysis indicates a positive correlation between intratumoural *Ncaph* levels and post-chemotherapy lung metastases, with higher *Ncaph* levels linked to increased metastases (*p*-values < .05). A positive 'Estimate' indicates a correlation between higher Ncaph levels and an increased count of lung metastases. (K and L) Poisson GLM analysis further confirms this correlation in mice treated with docetaxel (K) and doxorubicin (L), where individual mouse data points display the relationship between *Ncaph* RNA levels and metastasis count. The model's red line illustrates the trend, underscoring *Ncaph's* role in lung metastasis after breast tumour treatment with these drugs.
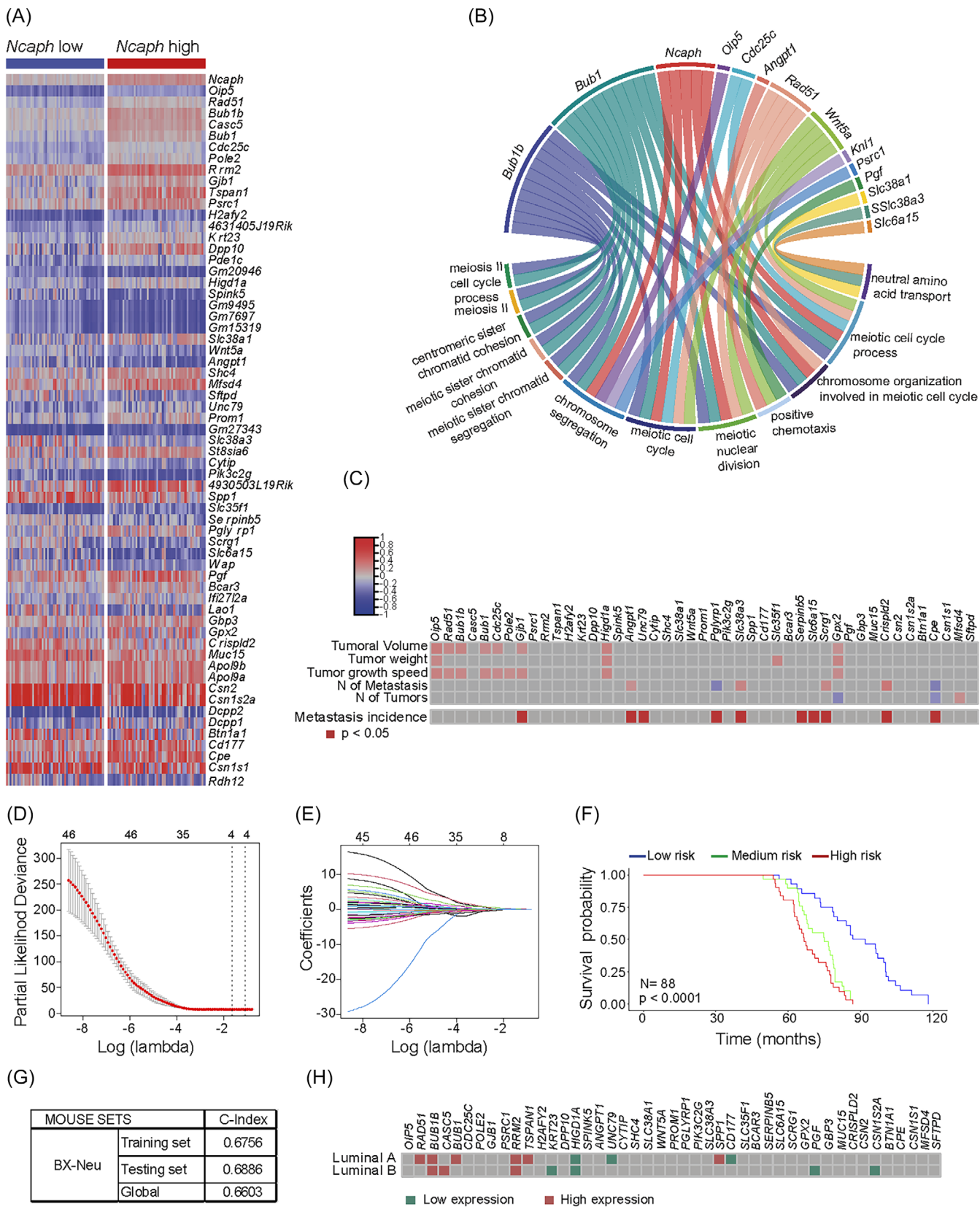
FIGURE 5 Identifying a gene signature associated with non-SMC condensin I complex subunit H (*NCAPH*) in the Mouse Mammary Tumor Virus (MMTV)-*ErbB2* cohort of mice generated by backcrossing (BX-*Neu*+) recognises the poor outcome in mice and humans. (A) Heatmap showing the 64 differentially expressed genes between tumours with high or low intratumoural levels of *Ncaph* in BX-*Neu*+ mice, of which 45 were shared with humans. Extreme tertiles of *Ncaph* expression were taken to define the high and low levels of *Ncaph*. The criteria

mitosis (Figures 5B and S5C and Table S4), and some were also correlated with poor breast cancer outcome in the BX-*Neu*+ cohort of mice (Figure 5C). When we integrated several of these genes into a multivariate LASSO regression model to define poor tumour outcome in the BX-*Neu*+ mouse cohort (Figure 5D,E and Table S5), four genes were identified that were associated with poor survival in BX-*Neu*+ mice: *Oip5*, *Higd1a*, *Shc4* and *Scrg1* (Figure S5D–F). Interestingly, the intratumoural levels of certain genes identified in the heterogeneous BX-*Neu*+ model, and associated with elevated levels of *Ncaph*, also correlated with adverse clinical outcomes (RFS) in patients with the intrinsic luminal A subtype of breast cancer[59] (Figure 5H and Table S6).

In conclusion, elevated expression of *NCAPH* was linked to a set of gene transcripts. The intratumoural levels of these transcripts, akin to *NCAPH* itself, were also associated with the adverse progression of luminal breast cancer in both mice and humans.

3.6 | Identification of a genetic model associated with poor outcome in patients with luminal tumours

Despite having the best overall prognosis among the intrinsic subtypes, luminal A tumours display significant variation in prognosis. It is crucial to identify patients with poor prognoses for improved survival via initial therapeutic enhancements. High levels of *NCAPH* were associated with poor outcome, especially in luminal A tumours (Figure 1D). Our study identified an array of genes that exhibited a notable correlation with elevated intratumoural levels of *Ncaph*, as depicted in Figure 5. We also found that some of these genes were associated with

poor RFS in humans (Figure 5H). Therefore, we used a penalised multivariate LASSO regression model to identify a gene signature that reflects the poor prognosis of luminal A tumours.[31]

The LASSO regression model was generated from the GOBO database of 401 patients with luminal A-diagnosed breast cancers.[41] This cohort was divided into a training set (70%) and a test set comprising the remaining 30% to generate a polygenic risk score. First, bivariate analyses of the training set using Cox regression identified genes associated with poor outcome in terms of RFS (Table S7). Later, genes with a *p*-value <.25 were used to generate a polygenic risk score using the restrictive LASSO regression model.[31] Thus, the LASSO model was developed, and ten genes were identified that define disease prognosis, which we referred to as the GSLA10 (Figures 6A and S6A,B and Table S8). The prognoses in the training and testing sets and the global model were evaluated using the *C*-index (Figure S6C,D).

GSLA10 discriminated between low-risk (bottom 1/3 risk score), medium-risk (middle 1/3 risk score) and high-risk (top 1/3 risk score) RFS in patients (Figure 6B). We validated our model in two independent patient cohorts, METABRIC and TCGA-BRCA, with 718 and 499 luminal A breast cancer cases, respectively. The *C*-index, area under the curve (AUC) of the ROC curve, and log-rank *p*-value of the KM analysis were assessed when applying GSLA10 to the METABRIC and TCGA-BRCA cohorts, confirming the predictive capability of GSLA10 (Figures 6C,D and S6D).

Additionally, we compared the ability of these signatures to define the prognosis of luminal A tumours in terms of RFS at different time points, evaluating the AUC for the first five years after diagnosis, between 5 and 10 years, and after 10 years in METABRIC and TCGA-BRCA (Figure 6E–G). The ROC curves indicated that GSLA10 can

for underexpressed and overexpressed genes were a ← twofold and a > twofold change, respectively, and a value of *p* < .05 (Table S3). (B) Gene Ontology analysis shows the main biological functions in which the 45 genes of the signature participate (Table S4). (C) Some of the genes in the 45-gene signature were associated with the poor outcome of breast cancer in MMTV-*ErbB2* mice generated by backcrossing. Correlation of the intratumoural levels of these genes with specific pathophenotypes of breast cancer in BX-*Neu*+ mice: Pearson's test. The incidence of metastasis was evaluated with a chi-squared test. (D–G) Least absolute shrinkage and selection operator (LASSO) regression model that defines poor outcome in BX-*Neu*+ mice. The regression coefficient map of genes in the LASSO model, using cross-validation to select the optimal tuning parameter (λ). The dotted vertical lines were drawn at the optimal values using the minimum criteria and 1 standard error (SE) of the minimum criteria (the 1 – SE criteria) (D). The graph shows the screening path of the LASSO regression model. Each curve represents a LASSO coefficient of the 45 prognostic genes, and the x-axis indicates the regularisation penalty parameter. When the number of variables was 4, the partial likelihood deviation was at the minimum, corresponding to the minimum λ value (E). The LASSO regression model generated differentiates between low-, intermediate- and high-risk BX-*Neu*+ mice in terms of survival probability and as defined by tertiles. Kaplan–Meier curve and log-rank test (F). Goodness-of-fit measures for the generated LASSO models in the BX-*Neu*+ mouse cohort. Training, testing and global model results (G). (H) The heatmap represents the association of gene transcript levels with the prognosis of human luminal breast tumours, in relation to *Ncaph* levels in mouse breast tumours from the backcross cohort. Colour coding: Brown indicates genes where high transcript levels are statistically significantly associated with poor prognosis; green denotes genes where low transcript levels correlate with poor prognosis. The heatmap is supported by detailed data on patient count, expression values, and *p*-values in Table S6. SMC, structural chromosome maintenance.

(A)

LASSO Model

Gene	Coefficient
Ncaph	0.0542
Bub1	0.1304
Rrm2	0.3973
Krt23	-0.0640
Higd1a	-0.4648
Sftpd	-0.0464
Slc38a3	0.1666
Pik3c2g	-0.2886
Gpx2	0.2346
Cpe	0.3303
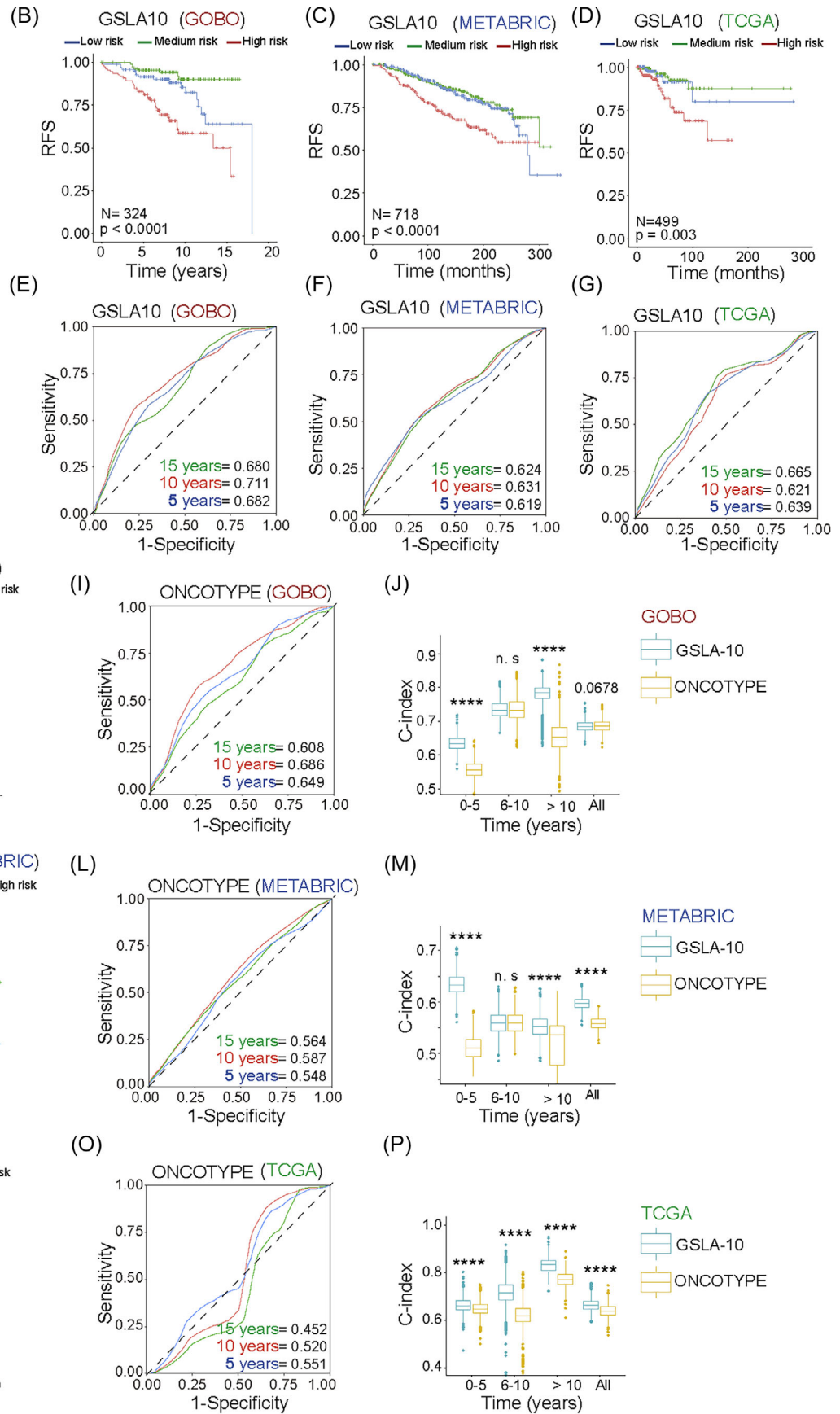
predict the risk of relapse in these patients at different time points.

In addition, we constructed a risk model (Oncotype) based on 21 genes in the Oncotype. Each Oncotype gene's expression level was adjusted in relation to a set of five reference genes, as detailed in Paik et al.[6] We provided a comprehensive comparison between GSLA10 and the Oncotype in terms of patient risk stratification and prognostic power in both the training cohort (GOBO) (Figure 6H–J) and the independent validation cohorts (METABRIC and TCGA-BRCA) (Figures 6K–M,N–P and S6E). The comparison further confirmed the robustness and superior prognostic power of GSLA10 over Oncotype in luminal A tumours.

In conclusion, the GSLA10 signature was associated with poor prognosis in luminal A patients. This model could help assess the prognosis of luminal A tumours and thus favor more personalised follow-up and therapy for patients with breast cancer (Figure 7).

4 | DISCUSSION

In our study, we discovered a gene signature (GSLA10) linked to high intratumoural NCAPH levels and the unfavourable progression of luminal A breast tumours. The need for accurate identification of the prognosis of this tumour type is imperative because of differing initial treatment responses, including the potential inclusion of chemotherapy.[70–73] Gene signatures, notably Oncotype DX, have been employed to identify potential chemotherapy beneficiaries among patients with ER-positive tumours.[6,74–76] Moreover, luminal A tumours

exhibit the highest post-10-year relapse risk despite endocrine therapy, necessitating precise patient identification and potential extended hormonal treatment.[77–79] Consequently, enhancing the prognostic precision of luminal A tumours is of crucial importance. In this context, our GSLA10 gene signature has shown superior prognostic power over Oncotype DX in both short- and long-term scenarios, suggesting its potential for tailoring luminal A patient treatment. Although GSLA10 demonstrated increased efficacy in predicting luminal A tumour prognosis, further studies are required to confirm whether this signature can reliably identify patients who may benefit from chemotherapy at diagnosis or from prolonged hormonal treatment.

We found that the overexpression of *NCAPH* is associated with poor prognosis, specifically in luminal A tumours and HER2$^+$ luminal tumours. The specific association between high NCAPH levels and poor prognosis in luminal tumours may be related to the ability of condensin I complex to bind to ER enhancers. Indeed, condensins play an essential role in activating the expression of estrogen target genes through their activity at the transcriptional level.[80] Moreover, the putative potentiation of estrogen signalling by NCAPH helps explain the hyperplasia observed in transgenic mice overexpressing NCAPH and their increased susceptibility to breast cancer development. This could also explain why NCAPH levels are not associated with the poor prognosis of breast tumours that do not express ER, such as HER2-enriched and basal tumours, as well as the participation of NCAPH in the poor clinical outcome of other hormone-dependent tumours, such as ovarian, endometrial, cervical and possibly prostate cancers.[81–86] Furthermore, luminal A tumours

FIGURE 6 Comparative evaluation of Gene Signature for Luminal A 10 (GSLA10) and the Oncotype signature for prognostic determination in luminal A tumours. Study and validation of GSLA10 in patients on the Gene Expression-based Outcome for Breast Cancer Online (GOBO), Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) and The Cancer Genome Atlas (TCGA) datasets. (A) Hazard ratio illustration of each GSLA10 gene's individual risk contribution, as implicated in the least absolute shrinkage and selection operator (LASSO) model from the GOBO cohort. (B) Assessment of the GSLA10 model's capability in delineating luminal A tumour progression via Kaplan–Meier curves from the GOBO cohort. (C) Validation of the robustness of the GSLA10 model for luminal A tumour progression definition using Kaplan–Meier curves in the METABRIC cohort. (D) Validation of the Oncotype model's robustness in characterising luminal A tumour progression with Kaplan–Meier curves from the METABRIC cohort. (E) Receiver operating characteristic (ROC) curves illustrating the prognostic prediction prowess of the GSLA10 model within the GOBO cohort. (F) ROC curves demonstrating the GSLA10 model's prognostic prediction in the METABRIC cohort. (G) ROC curves of the GSLA10 model for prognosis prediction in the METABRIC cohort. (H) The Oncotype model's efficacy in defining luminal A tumour outcome was evaluated with Kaplan–Meier curves in the GOBO cohort. (I) ROC curves showcasing the Oncotype model's ability to predict prognosis within the GOBO cohort. (J) C-index comparison for prognostic prediction between GSLA10 and Oncotype models within the GOBO cohort. (K) Validation of the Oncotype model's robustness in characterising luminal A tumour progression using Kaplan–Meier curves in the METABRIC cohort. (L) ROC curves of the Oncotype model for prognostic prediction in the METABRIC cohort. (M) Comparison of the C-index for prognostic prediction between the GSLA10 and Oncotype models in the METABRIC cohort. (N) Further validation of the Oncotype model's robustness in defining luminal A tumour progression, as illustrated by Kaplan–Meier curves from the METABRIC cohort. (O) ROC curves of the Oncotype model for prognostic prediction in the METABRIC cohort. (P) Comparative evaluation of the C-index for prognostic prediction between the GSLA10 and Oncotype models in the METABRIC cohort. Each graph in the Kaplan–Meier curve panels displays the number of patients it encompasses.
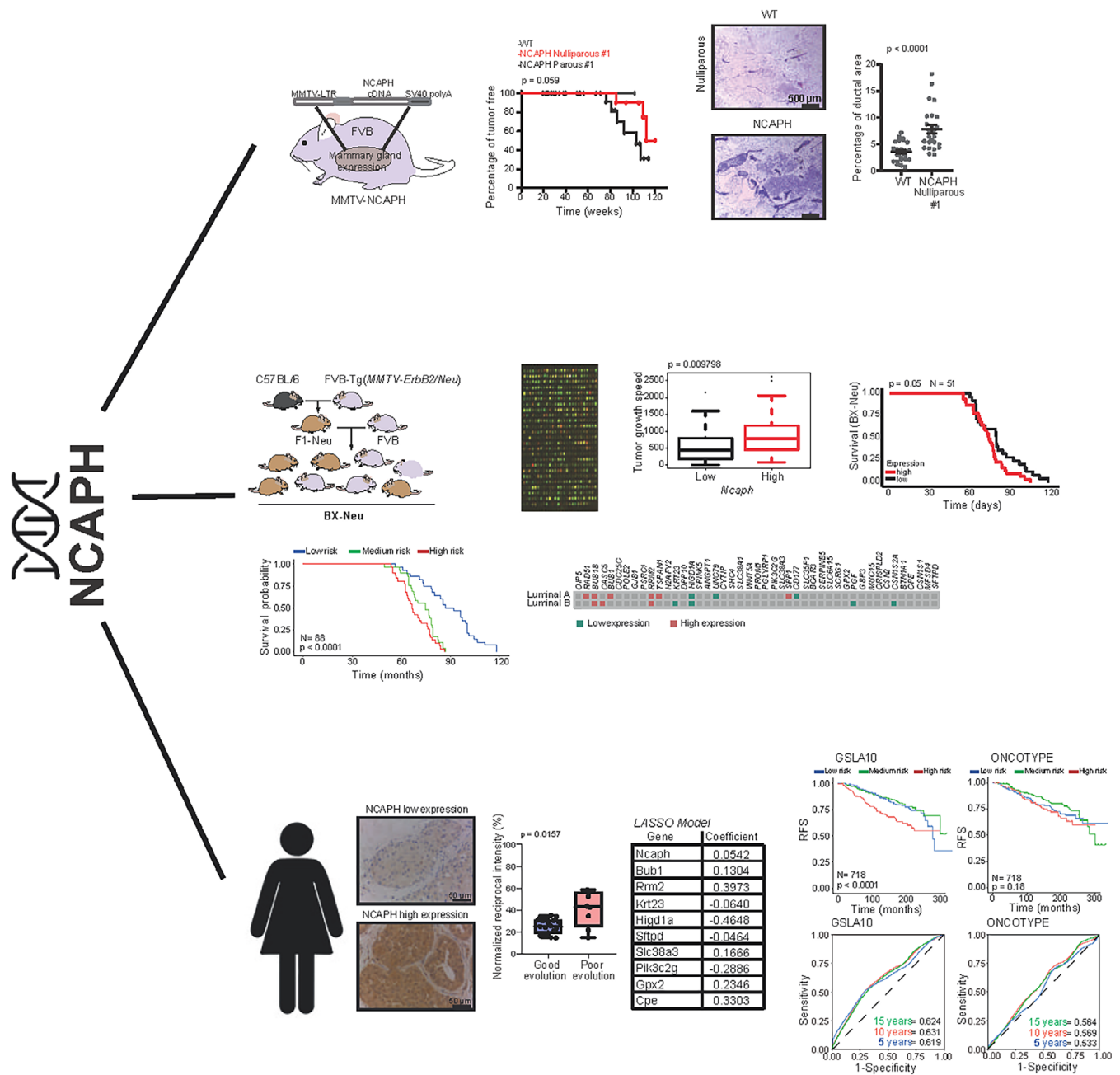
FIGURE 7 Graphical representation of non-SMC condensin I complex subunit H's (NCAPH's) role in luminal A breast cancer progression. Using transgenic mouse models and a genetically diverse backcrossed mouse cohort, elevated NCAPH expression was linked to unfavourable tumour development. A heatmap showcases intratumoural NCAPH expression, correlating with poor prognosis. Least absolute shrinkage and selection operator (LASSO) analysis identified a 10-gene signature, Gene Signature for Luminal A 10 (GSLA10), presented alongside its superior predictive capability compared to Oncotype DX, suggesting its potential in guiding personalised luminal A breast cancer treatments. SMC, structural chromosome maintenance.

with worse outcome are thought to have higher genomic instability, alterations to P53[19] and overexpression of genes that regulate mitosis.[20]

In the last 5 years, high levels of NCAPH expression have been associated with the pathogenesis and prognosis of several tumour types,[81,82,84,86–91] including hepatocarcinoma,[92] lung tumours,[93–95] melanoma[96] and endometrial cancer.[83] During the development of this project, two studies on NCAPH in breast cancer were published. The first study demonstrated that in vitro, NCAPH expression is elevated in MCF7 cells compared to the non-tumourigenic breast cell line, MCF10A.[22] The second study revealed that downregulating NCAPH in MCF7 cells leads to a reduction in proliferation.[47] In the present work,

we substantially broaden the research on NCAPH's role in breast cancer development. In vitro, we have demonstrated its involvement in viability, proliferation, increased genomic instability, alterations in cellular signalling at various levels and resistance to multiple treatment modalities. Furthermore, we have introduced an in vivo transgenic mouse model that overexpresses NCAPH for the first time. Our findings suggest that NCAPH overexpression can eventually act as a primary oncogenic trigger for breast cancer. Notably, the overexpression of NCAPH leads to an enlargement of the non-tumourous breast's glandular component, a recognised factor predisposing to breast cancer.[97]

Paradoxically, elevated NCAPH levels were associated with good evolution of basal tumours. Notably, NCAPH depletion is associated with inhibition of proliferation, migration and xenograft tumour formation in colon cancer cell lines. However, elevated NCAPH levels in patients with tumours have been associated with a better prognosis and survival rate than in patients with low levels of NCAPH.[89] Similar results have been reported for human papillomavirus-positive cervical cancer.[86] For both tumour types, these features have been related to the enhanced proliferation induced by NCAPH, which might sensitise tumour cells to chemotherapy and radiation therapy.[86,89]

Higher NCAPH levels linked to better outcomes in some cancers might potentially be due to their role in triggering oncogene-induced apoptosis or senescence, a phenomenon also observed in other mitosis-regulating genes such as Aurora A, Aurora B, Polo-Kinase-1, Cyc E and CDC25.[98–101] These genes' effects can vary, leading to either positive or negative outcomes, depending on how they interact with other oncogenic changes, such as p53 status.[101] Moreover, increased cell proliferation and apoptosis have been noted with the overexpression of certain oncogenes like *MYC*, *E2A* and *E2F1*. Blocking apoptosis or senescence pathways in these cases allows cells to better tolerate high molecular levels and the resulting increased proliferation.[102–104] While further investigation into the tolerance of elevated NCAPH levels and its interaction with other secondary oncogenic events is beyond the scope of our current study, the increased apoptosis or senescence triggered by NCAPH, due to enhanced proliferation and genomic stress, might explain our findings. Specifically, we observed a marginally reduced tumour incidence in instances where NCAPH is induced as a primary oncogenic event in normal breast tissue, as compared to non-induced cases. This was evident in MMTV-*Ncaph* multiparous mice in contrast to their nulliparous counterparts. Additionally, a higher rate of apoptosis was noted in MMTV-*Ncaph*[ErbB2] double-transgenic tumours than in single-transgenic MMTV-*ErbB2* tumours. In these cases, although higher levels of *Ncaph* were associated with

increased tumour proliferation, the apoptosis (and potentially senescence) induced by Ncaph and the oncogenic stress it causes were still observable.

Our study also reveals a correlation between NCAPH and poor outcomes in HER2-positive luminal tumours in both humans and mice. In mice, this association results in breasts having an expanded glandular component and heightened proliferation both of which are recognised as breast cancer risk factors.[105–107] It was evident that HER2[+] luminal tumours overexpressing *NCAPH* are more aggressive, characterised by rapid growth, increased proliferation and a high mitotic index. Collectively, our discoveries substantially augment the current understanding of NCAPH's role in the onset and progression of breast cancer.

Given our findings that elevated NCAPH levels correlate with poorer outcomes in luminal ERBB2 tumours in transgenic mice, we identified a gene signature associated with high NCAPH levels. This was also conducted using a backcross cohort of mice with luminal ERBB2 tumours, as they epitomise an extension of the phenotypic presentation of breast cancer and its associated transcriptomics.[50] In backcross models, a notable phenotypic variation in breast cancer is observed alongside a heightened transcriptomic variation, facilitating the association of both phenotype and transcriptome with their genetic regulation.[66,64,69] Genetically heterogeneous mouse cohorts better reflect the heterogeneity observed in the human population, facilitating the identification of the genetic and transcriptomic determinants associated with disease outcome.[35,66,67] In this study, we utilised genetically heterogeneous mouse cohorts as a model system. This approach is particularly relevant because it mirrors the genetic diversity inherent in the human population. Such heterogeneity is a crucial factor in biomedical research, as it allows for a more accurate representation of the complexities and variabilities observed in human diseases. This modelling is especially significant in the context of identifying genetic and transcriptomic determinants that drive disease outcome. Studies have shown that the use of genetically uniform models often fails to capture the full spectrum of disease manifestations.[35,66,67] In contrast, heterogeneous cohorts provide a broader genetic context, which is instrumental in uncovering the multifaceted nature of gene–disease interactions. This diversity enables the identification of subtler genetic variants and transcriptomic profiles that might be overlooked in more genetically homogeneous models. Enrichment analyses revealed that some of the identified genes govern cell cycle progression and mitosis.

Interestingly, some of these genes were also associated with poor breast cancer outcome in BX-*Neu*[+] mice, both individually and after the application of the LASSO regression model. It is important to note that some genes in this signature were individually associated with poor out-

come of intrinsic luminal A subtypes in humans in the KM plotter database.[59] Moreover, the application of some of these genes in a multivariate LASSO regression model identified luminal A tumours that relapsed in three human breast cancer datasets. It is plausible that patients with luminal A tumours, exhibiting high levels of NCAPH and displaying the associated signature could represent a distinct subgroup of tumours. This subgroup might exhibit an intermediate prognosis, positioned between those observed for the luminal A and luminal B subtypes. Further research is required to investigate this hypothesis and elucidate its potential taxonomic implications.

Our encouraging findings with GSLA10 indicate the potential value of conducting future prospective studies. These studies should assess the efficacy of GSLA10 both independently and in combination with other signature biomarkers like OncotypeDX, specifically evaluating their impact on therapeutic decision making and patient outcomes.

5 | CONCLUSIONS

In conclusion, we have demonstrated that NCAPH is involved in the pathogenesis of breast cancer. Moreover, an NCAPH-associated signature defines the outcome of the luminal A breast cancer subtype. The potential of GSLA10 to distinguish a specific cohort of patients with luminal A tumours, who might benefit significantly from intensified treatment protocols aimed at preventing relapse, is indeed stimulating. Certainly, the GSLA10 signature could serve as a diagnostic tool to identify patients with luminal A tumours who are at risk of poor prognosis. This insight may enable healthcare providers to devise personalised treatment strategies that are more efficacious for these patients.

AUTHOR CONTRIBUTIONS

Marina Mendiburu-Eliçabe contributed to the study, generation of LASSO models and statistical analyses of the mouse and patient databases. He also critically reviewed the manuscript. Natalia García-Sancha and Roberto Corchado-Cobos have been used in studies involving transgenic mice, including generation of mouse lines, necropsies, tissue analysis, histopathology and IHC. Both provided a critical review of the manuscript. Angélica Martínez-López studies have been conducted on cell lines, encompassing the generation of stable lines, cultures, feasibility analysis, proliferation and signalling pathways. He critically reviewed the manuscript. Hang Chang and Jian Hua Mao engaged in human database studies comparing GSLA10 with Oncotype DX and supervised the statistical analyses of the manuscript. Adrián Blanco-Gómez

analysed mice produced by backcrossing and studied allogeneic transplantation and chemotherapy treatments. Ana García-Casas, Andrés Castellanos-Martín and Nélida Salvador contributed to in vitro studies, generation of stable cell lines and signalling pathway analysis. Andrés Castellanos-Martín specifically focuses on the generation and analysis of mice produced by backcrossing. Alejandro Jiménez-Navas was involved in revising the survival curves and quantifying human IHC studies using software. Manuel Jesús Pérez-Baena participated in in vitro studies on stable triple-negative breast cancer cells. Manuel Adolfo Sánchez-Martín contributed to the generation of transgenic mice and microinjection and chimera analyses. María Del Mar Abad-Hernández and Sofía Del Carmen, both pathologists, evaluated the histopathology and IHC samples of luminal A human breast tumours. Juncal Claros-Ampuero, Juan Jesús Cruz-Hernández and César Augusto Rodríguez-Sánchez are oncologists who contributed to the diagnosis and clinical follow-up of breast cancer patients, and their tumours were analysed using IHC in this study. María Begoña García-Cenador and Francisco Javier García-Criado provided logistic support, focusing on animal maintenance and development of surgical protocols for breast tumour interventions. Rodrigo Santamaría Vicente conducted the preliminary bioinformatics studies. Sonia Castillo-Lluva and Jesús Pérez-Losada played pivotal roles in the project's conception, supervision and administration. They secured funding, managed the project and were actively involved in writing, editing and critical review of the manuscript. Their contributions also spanned the study and analysis of transgenic mice, and the conceptualisation and revision of statistical data.

CONFLICT OF INTEREST STATEMENT
The authors declare they have no conflicts of interest.

DATA AVAILABILITY STATEMENT
All data generated or analysed during this study are included in this published article (and its supporting information files). The datasets generated and/or analysed during the current study are available from the corresponding authors upon reasonable request. The datasets analysed during the current study are available from GOBO (http://co.bmc.lu.se/gobo),[41] METABRIC (https://www.cbioportal.org/) and TCGA-BRCA (https://www.cbioportal.org/).

ETHICS STATEMENT
All the mice were housed at the Animal Research Facility of the University of Salamanca. All procedures were approved by the Institutional Animal Care and Bioethics Committee of the University of Salamanca. Committee reference number: 197-2017. Human primary breast tumours were collected at the University Hospital of Salamanca (Salamanca, Spain), after the hospital's Institutional Ethics Review Board approved the protocols for collecting and using patient samples. Written informed consent was obtained from all patients to conduct the study on these tumour samples. Committee reference number: PI 2019 12 396.

CONSENT FOR PUBLICATION
The Consent for Publication statement is satisfactory and in compliance with the required ethical standards. This assurance is based on the fact that our study does not include any photographs or data that could potentially reveal the identity of the participants involved. We have taken diligent care to ensure that all personal information has been appropriately anonymized to maintain the confidentiality and privacy of the subjects.

ORCID
Natalia García-Sancha https://orcid.org/0000-0001-7731-7020
Hang Chang https://orcid.org/0000-0002-3773-6818
Jian Hua Mao https://orcid.org/0000-0001-9320-6021
Alejandro Jiménez-Navas https://orcid.org/0000-0001-7621-576X
Sonia Castillo-Lluva https://orcid.org/0000-0001-5357-7178
Jesús Pérez-Losada https://orcid.org/0000-0003-2400-624X

REFERENCES
1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-249.
2. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415(6871):530-536.
3. Buyse M, Loi S, van't Veer L, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst*. 2006;98(17):1183-1192.
4. Cardoso F, Piccart-Gebhart M, Van't Veer L, Rutgers E, Consortium T. The MINDACT trial: the first prospective clinical validation of a genomic tool. *Mol Oncol*. 2007;1(3):246-251.
5. Nielsen TO, Parker JS, Leung S, et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clin Cancer Res*. 2010;16(21):5222-5232.
6. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004;351(27):2817-2826.
7. Prat A, Ellis MJ, Perou CM. Practical implications of gene-expression-based assays for breast oncologists. *Nat Rev Clin Oncol*. 2011;9(1):48-57.
8. Tang G, Shak S, Paik S, et al. Comparison of the prognostic and predictive utilities of the 21-gene recurrence score assay and adjuvant! for women with node-negative, ER-positive breast cancer: results from NSABP B-14 and NSABP B-20. *Breast Cancer Res Treat*. 2011;127(1):133-142.
9. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009;27(8):1160-1167.
10. Ellis MJ, Suman VJ, Hoog J, et al. Randomized phase II neoadjuvant comparison between letrozole, anastrozole, and exemestane for postmenopausal women with estrogen receptor-rich stage 2 to 3 breast cancer: clinical and biomarker outcomes and predictive value of the baseline PAM50-based intrinsic subtype–ACOSOG Z1031. *J Clin Oncol*. 2011;29(17):2342-2349.
11. Sestak I, Buus R, Cuzick J, et al. Comparison of the performance of 6 prognostic signatures for estrogen receptor-positive breast cancer: a secondary analysis of a randomized clinical trial. *JAMA Oncol*. 2018;4(4):545-553.
12. Sestak I, Cuzick J. Markers for the identification of late breast cancer recurrence. *Breast Cancer Res*. 2015;17:10.
13. Ignatov A, Eggemann H, Burger E, Ignatov T. Patterns of breast cancer relapse in accordance to biological subtype. *J Cancer Res Clin Oncol*. 2018;144(7):1347-1355.

14. Pedersen RN, Esen BO, Mellemkjaer L, et al. The incidence of breast cancer recurrence 10–32 years after primary diagnosis. *J Natl Cancer Inst*. 2022;114(3):391-399.

15. Natarajan L, Pu M, Parker BA, et al. Time-varying effects of prognostic factors associated with disease-free survival in breast cancer. *Am J Epidemiol*. 2009;169(12):1463-1470.

16. Colleoni M, Sun Z, Price KN, et al. Annual hazard rates of recurrence for breast cancer during 24 years of follow-up: results from the International Breast Cancer Study Group Trials I to V. *J Clin Oncol*. 2016;34(9):927-935.

17. Ogba N, Manning NG, Bliesner BS, et al. Luminal breast cancer metastases and tumor arousal from dormancy are promoted by direct actions of estradiol and progesterone on the malignant cells. *Breast Cancer Res*. 2014;16(6):489.

18. Pu M, Messer K, Davies SR, et al. Research-based PAM50 signature and long-term breast cancer survival. *Breast Cancer Res Treat*. 2020;179(1):197-206.

19. Ciriello G, Sinha R, Hoadley KA, et al. The molecular diversity of luminal A breast tumors. *Breast Cancer Res Treat*. 2013;141(3):409-420.

20. Perez-Pena J, Alcaraz-Sanabria A, Nieto-Jimenez C, et al. Mitotic read-out genes confer poor outcome in luminal A breast cancer tumors. *Oncotarget*. 2017;8(13):21733-21740.

21. Hirano T. Condensins: universal organizers of chromosomes with diverse functions. *Genes Dev*. 2012;26(15):1659-1678.

22. Ono T, Losada A, Hirano M, Myers MP, Neuwald AF, Hirano T. Differential contributions of condensin I and condensin II to mitotic chromosome architecture in vertebrate cells. *Cell*. 2003;115(1):109-121.

23. Hirota T, Gerlich D, Koch B, Ellenberg J, Peters JM. Distinct functions of condensin I and II in mitotic chromosome assembly. *J Cell Sci*. 2004;117(pt 26):6435-6445.

24. Ono T, Fang Y, Spector DL, Hirano T. Spatial and temporal regulation of condensins I and II in mitotic chromosome assembly in human cells. *Mol Biol Cell*. 2004;15(7):3296-3308.

25. Maeshima K, Laemmli UK. A two-step scaffolding model for mitotic chromosome assembly. *Dev Cell*. 2003;4(4):467-480.

26. Gerlich D, Hirota T, Koch B, Peters JM, Ellenberg J. Condensin I stabilizes chromosomes mechanically through a dynamic interaction in live cells. *Curr Biol*. 2006;16(4):333-344.

27. Neuwald AF, Hirano T. HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions. *Genome Res*. 2000;10(10):1445-1452.

28. Hirano T. Condensins: organizing and segregating the genome. *Curr Biol*. 2005;15(7):R265-R275.

29. Hirano T. Condensin-based chromosome organization from bacteria to vertebrates. *Cell*. 2016;164(5):847-857.

30. Weyburne E, Bosco G. Cancer-associated mutations in the condensin II subunit CAPH2 cause genomic instability through telomere dysfunction and anaphase chromosome bridges. *J Cell Physiol*. 2021;236(5):3579-3598.

31. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med*. 1997;16(4):385-395.

32. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012;9(7):676-682.

33. Nguyen D, Quantifying chromogen intensity in immuno-histochemistry via reciprocal intensity. *CancerInCytes*. 2013. www.cancerincytes.org/quantifying-chromogen-intensity-in-immunohistochemistry-#!

34. Guy CT, Webster MA, Schaller M, Parsons TJ, Cardiff RD, Muller WJ. Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease. *Proc Natl Acad Sci U S A*. 1992;89(22):10578-10582.

35. Castellanos-Martin A, Castillo-Lluva S, Saez-Freire Mdel M, et al. Unraveling heterogeneous susceptibility and the evolution of breast cancer using a systems biology approach. *Genome Biol*. 2015;16:40.

36. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25(1):25-29.

37. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics*. 2012;16(5):284-287.

38. Therneau T. *A Package for Survival Analysis in R*. R package version 3.4-0. 2022. https://CRAN.R-project.org/package=survival

39. Therneau T, Grambsch P. *Modeling Survival Data: Extending the Cox Model*. Springer; 2000.

40. Kassambara A, Kosinski M, Biecek P, Fabian S. *Package 'survminer'. Drawing Survival Curves using 'ggplot2' (R package version 03 1)*. 2017.

41. Ringner M, Fredlund E, Hakkinen J, Borg A, Staaf J. GOBO: gene expression-based outcome for breast cancer online. *PLoS One*. 2011;6(3):e17911.

42. Hosmer DW Jr, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. John Wiley & Sons; 2013.

43. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer; 2001.

44. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinf*. 2011;12(1):1-8.

45. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;486(7403):346-352.

46. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61-70.

47. Lawrence I, Lin K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989;45(1):255-268.

48. Longato E, Vettoretti M, Di Camillo B. A practical perspective on the concordance index for the evaluation and selection of prognostic time-to-event models. *J Biomed Inform*. 2020;108:103496.

49. Castillo S, Blanco-Gómez A, Pérez-Losada J. Papel de la Kleisina NCAPH en el desarrollo y evolución del cáncer de mama HER2 positivo. *XXXIX Congress of the Spanish Society of Biochemistry and Molecular Biology (SEBBM)*. 2016.

50. García-Sancha N, Corchado-Cobos R, Martínez-López A, et al. Analyzing the role of non-SMC condensin I complex subunit H in breast cancer development and evolution. *Symposium on Cancer Genomics and Epitranscriptomics: From the Bench to the Clinic*. 2022.

51. Riis ML, Zhao X, Kaveh F, et al. Gene expression profile analysis of t1 and t2 breast cancer reveals different activation pathways. *ISRN Oncol*. 2013;2013:924971.

52. Pawitan Y, Bjohle J, Amler L, et al. Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res.* 2005;7(6):R953-R964.

53. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet.* 2005;365(9460):671-679.

54. Ivshina AV, George J, Senko O, et al. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res.* 2006;66(21):10292-10301.

55. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med.* 2002;347(25):1999-2009.

56. Chandrashekar DS, Karthikeyan SK, Korla PK, et al. UAL-CAN: an update to the integrated cancer data analysis platform. *Neoplasia.* 2022;25:18-27.

57. Otten AD, Sanders MM, McKnight GS. The MMTV LTR promoter is induced by progesterone and dihydrotestosterone but not by estrogen. *Mol Endocrinol.* 1988;2(2):143-147.

58. Jezequel P, Gouraud W, Ben Azzouz F, et al. bc-GenExMiner 4.5: new mining module computes breast cancer differential gene expression analyses. *Database.* 2021;2021:baab007.

59. Lanczky A, Gyorffy B. Web-based survival analysis tool tailored for medical research (KMplot): development and implementation. *J Med Internet Res.* 2021;23(7):e27633.

60. Bastien RR, Rodriguez-Lescure A, Ebbert MT, et al. PAM50 breast cancer subtyping by RT-qPCR and concordance with standard clinical molecular markers. *BMC Med Genet.* 2012;5:44.

61. Vasconcelos I, Hussainzada A, Berger S, et al. The St. Gallen surrogate classification for breast cancer subtypes successfully predicts tumor presenting features, nodal involvement, recurrence patterns and disease free survival. *Breast.* 2016;29:181-185.

62. Cortazar AR, Torrano V, Martín-Martín N, et al. CANCER-TOOL: a visualization and representation interface to exploit cancer datasets. *Cancer Res.* 2018;78(21):6320-6328.

63. Saez-Freire MDM, Blanco-Gomez A, Castillo-Lluva S, et al. The biological age linked to oxidative stress modifies breast cancer aggressiveness. *Free Radical Biol Med.* 2018;120:133-146.

64. Schadt EE, Monks SA, Drake TA, et al. Genetics of gene expression surveyed in maize, mouse and man. *Nature.* 2003;422(6929):297-302.

65. Sanchez-Vega F, Mina M, Armenia J, et al. Oncogenic signaling pathways in The Cancer Genome Atlas. *Cell.* 2018;173(2):321-337.e10.

66. Balmain A. Cancer as a complex genetic trait: tumor susceptibility in humans and mouse models. *Cell.* 2002;108(2):145-152.

67. Mao JH, Perez-Losada J, Wu D, et al. Fbxw7/Cdc4 is a p53-dependent, haploinsufficient tumour suppressor gene. *Nature.* 2004;432(7018):775-779.

68. Auchincloss H Jr, Winn HJ. Clarence Cook Little (1888–1971): the genetic basis of transplant immunology. *Am J Transpl.* 2004;4(2):155-159.

69. Quigley DA, To MD, Perez-Losada J, et al. Genetic architecture of mouse skin inflammation and tumour susceptibility. *Nature.* 2009;458(7237):505-508.

70. Twelves C, Jove M, Gombos A, Awada A. Cytotoxic chemotherapy: still the mainstay of clinical practice for all subtypes metastatic breast cancer. *Crit Rev Oncol Hematol.* 2016;100:74-87.

71. Cardoso F, Kyriakides S, Ohno S, et al. Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-updagger. *Ann Oncol.* 2019;30(8):1194-1220.

72. Diessner J, Wischnewsky M, Blettner M, et al. Do patients with luminal A breast cancer profit from adjuvant systemic therapy? A retrospective multicenter study. *PLoS One.* 2016;11(12):e0168730.

73. Li Y, Ma L. Efficacy of chemotherapy for lymph node-positive luminal A subtype breast cancer patients: an updated meta-analysis. *World J Surg Oncol.* 2020;18(1):316.

74. Sparano JA, Gray RJ, Makower DF, et al. Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. *N Engl J Med.* 2018;379(2):111-121.

75. Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol.* 2006;24(23):3726-3734.

76. Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol.* 2023;41(20):3565-3575.

77. Pan H, Gray R, Braybrooke J, et al. 20-Year risks of breast-cancer recurrence after stopping endocrine therapy at 5 years. *N Engl J Med.* 2017;377(19):1836-1846.

78. Davies C, Pan H, Godwin J, et al. Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet.* 2013;381(9869):805-816.

79. Goss PE, Ingle JN, Pritchard KI, et al. Extending aromatase-inhibitor adjuvant therapy to 10 years. *N Engl J Med.* 2016;375(3):209-219.

80. Li W, Hu Y, Oh S, et al. Condensin I and II complexes license full estrogen receptor alpha-dependent enhancer activation. *Mol Cell.* 2015;59(2):188-202.

81. Zhan SJ, Liu B, Linghu H. Identifying genes as potential prognostic indicators in patients with serous ovarian cancer resistant to carboplatin using integrated bioinformatics analysis. *Oncol Rep.* 2018;39(6):2653-2663.

82. Qi Y, Mo K, Zhang T. A transcription factor that promotes proliferation, migration, invasion, and epithelial-mesenchymal transition of ovarian cancer cells and its possible mechanisms. *Biomed Eng Online.* 2021;20(1):83.

83. Qiu X, Gao Z, Shao J, Li H. NCAPH is upregulated in endometrial cancer and associated with poor clinicopathologic characteristics. *Ann Hum Genet.* 2020;84(6):437-446.

84. Cui F, Hu J, Xu Z, Tan J, Tang H. Overexpression of NCAPH is upregulated and predicts a poor prognosis in prostate cancer. *Oncol Lett.* 2019;17(6):5768-5776.

85. Arai T, Kojima S, Yamada Y, et al. Micro-ribonucleic acid expression signature of metastatic castration-resistant prostate cancer: regulation of NCAPH by antitumor miR-199a/b-3p. *Int J Urol.* 2019;26(4):506-520.

86. Wang M, Qiao X, Cooper T, et al. HPV E7-mediated NCAPH ectopic expression regulates the carcinogenesis of cervical carcinoma via PI3K/AKT/SGK pathway. *Cell Death Dis.* 2020;11(12):1049.

87. Xiong YC, Wang J, Cheng Y, Zhang XY, Ye XQ. Overexpression of MYBL2 promotes proliferation and migration of non-small-cell lung cancer via upregulating NCAPH. *Mol Cell Biochem*. 2020;468(1-2):185-193.

88. Shimomura H, Sasahira T, Nakashima C, Kurihara-Shimomura M, Kirita T. Non-SMC condensin I complex subunit H (NCAPH) is associated with lymphangiogenesis and drug resistance in oral squamous cell carcinoma. *J Clin Med*. 2019;9(1):72.

89. Yin L, Jiang LP, Shen QS, et al. NCAPH plays important roles in human colon cancer. *Cell Death Dis*. 2017;8(3):e2680.

90. Li B, Xiao Q, Shan L, Song Y. NCAPH promotes cell proliferation and inhibits cell apoptosis of bladder cancer cells through MEK/ERK signaling pathway. *Cell Cycle*. 2022;21(4):427-438.

91. Wang Y, Li JQ, Yang ZL, et al. NCAPH regulates gastric cancer progression through DNA damage response. *Neoplasma*. 2022;69(2):283-291.

92. Sun C, Huang S, Wang H, et al. Non-SMC condensin I complex subunit H enhances proliferation, migration, and invasion of hepatocellular carcinoma. *Mol Carcinog*. 2019;58(12):2266-2275.

93. Ma Q, Xu Y, Liao H, et al. Identification and validation of key genes associated with non-small-cell lung cancer. *J Cell Physiol*. 2019;234(12):22742-22752.

94. Zhou W, Hu J, Zhao J. Non-SMC condensin I complex subunit H (NCAPH), a regulator of cell cycle, predicts poor prognosis in lung adenocarcinoma patients: a study mainly based on TCGA and GEO database. *Transl Cancer Res*. 2020;9(12):7572-7587.

95. Li Z, Sang M, Tian Z, et al. Identification of key biomarkers and potential molecular mechanisms in lung cancer by bioinformatics analysis. *Oncol Lett*. 2019;18(5):4429-4440.

96. Ryu B, Kim DS, Deluca AM, Alani RM. Comprehensive expression profiling of tumor cell lines identifies molecular signatures of melanoma progression. *PLoS One*. 2007;2(7):e594.

97. Lee SH, Ryu HS, Jang MJ, et al. Glandular tissue component and breast cancer risk in mammographically dense breasts at screening breast US. *Radiology*. 2021;301(1):57-65.

98. Schmitt CA. Senescence, apoptosis and therapy—cutting the lifelines of cancer. *Nat Rev Cancer*. 2003;3(4):286-295.

99. Liu X-L, Ding J, Meng L-H. Oncogene-induced senescence: a double edged sword in cancer. *Acta Pharmacol Sin*. 2018;39(10):1553-1558.

100. Hills SA, Diffley JF. DNA replication and oncogene-induced replicative stress. *Curr Biol*. 2014;24(10):R435-R444.

101. Mao J-H, Wu D, Perez-Losada J, et al. Crosstalk between Aurora-A and p53: frequent deletion or downregulation of Aurora-A in tumors from p53 null mice. *Cancer Cell*. 2007;11(2):161-173.

102. Lowe SW, Cepero E, Evan G. Intrinsic tumour suppression. *Nature*. 2004;432(7015):307-315.

103. Murphy DJ, Junttila MR, Pouyet L, et al. Distinct thresholds govern Myc's biological output in vivo. *Cancer Cell*. 2008;14(6):447-457.

104. Larsson L-G, editor. Oncogene-and tumor suppressor gene-mediated suppression of cellular senescence. *Seminars in Cancer Biology*. Elsevier; 2011.

105. Nassar A, Hoskin TL, Stallings-Mann ML, et al. Ki-67 expression in sclerosing adenosis and adjacent normal breast terminal ductal lobular units: a nested case-control study from the Mayo Benign Breast Disease Cohort. *Breast Cancer Res Treat*. 2015;151(1):89-97.

106. Huh SJ, Oh H, Peterson MA, et al. The proliferative activity of mammary epithelial cells in normal tissue predicts breast cancer risk in premenopausal women. *Cancer Res*. 2016;76(7):1926-1934.

107. Oh H, Eliassen AH, Wang M, et al. Expression of estrogen receptor, progesterone receptor, and Ki67 in normal breast tissue in relation to subsequent risk of breast cancer. *NPJ Breast Cancer*. 2016;2:16032.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.