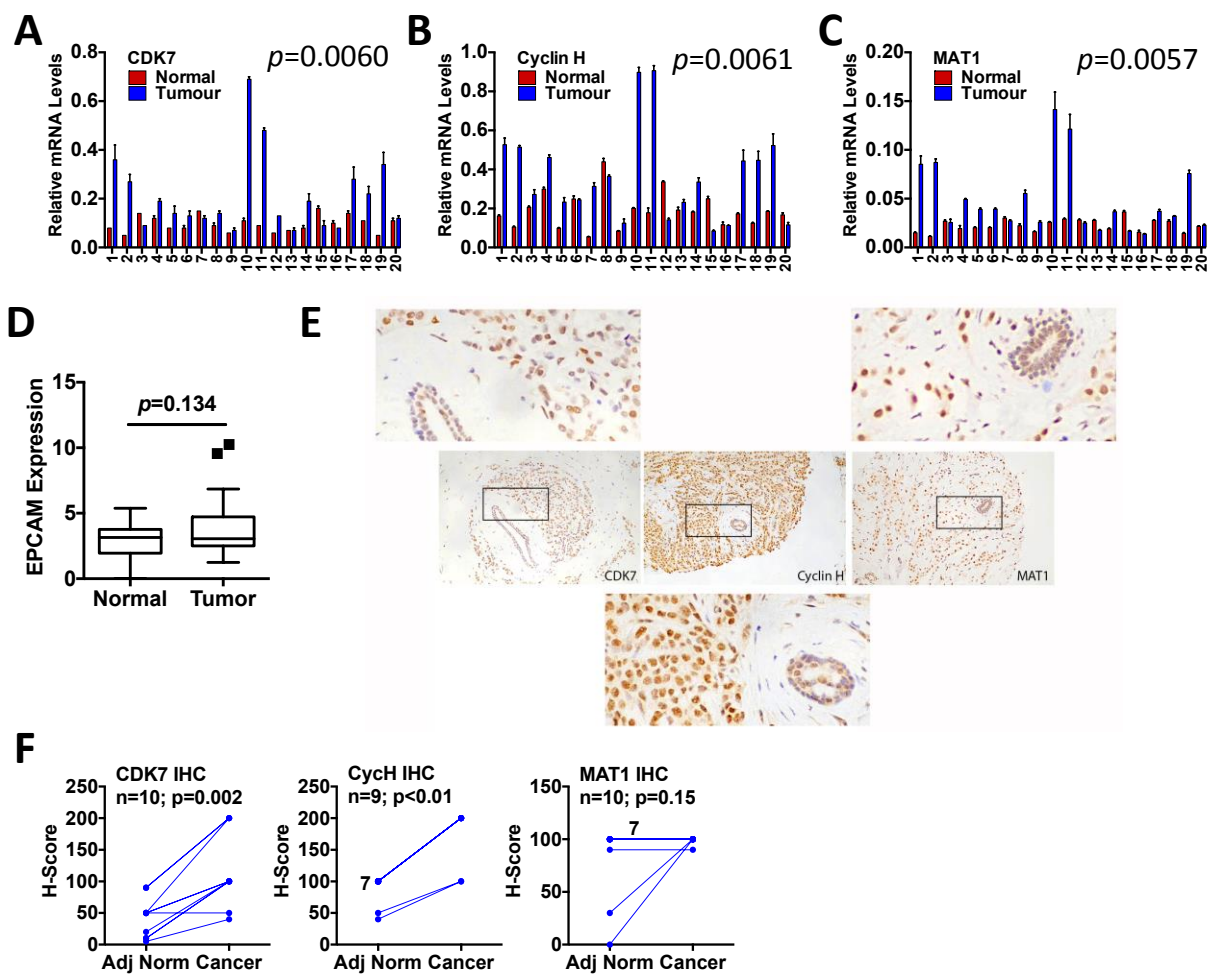


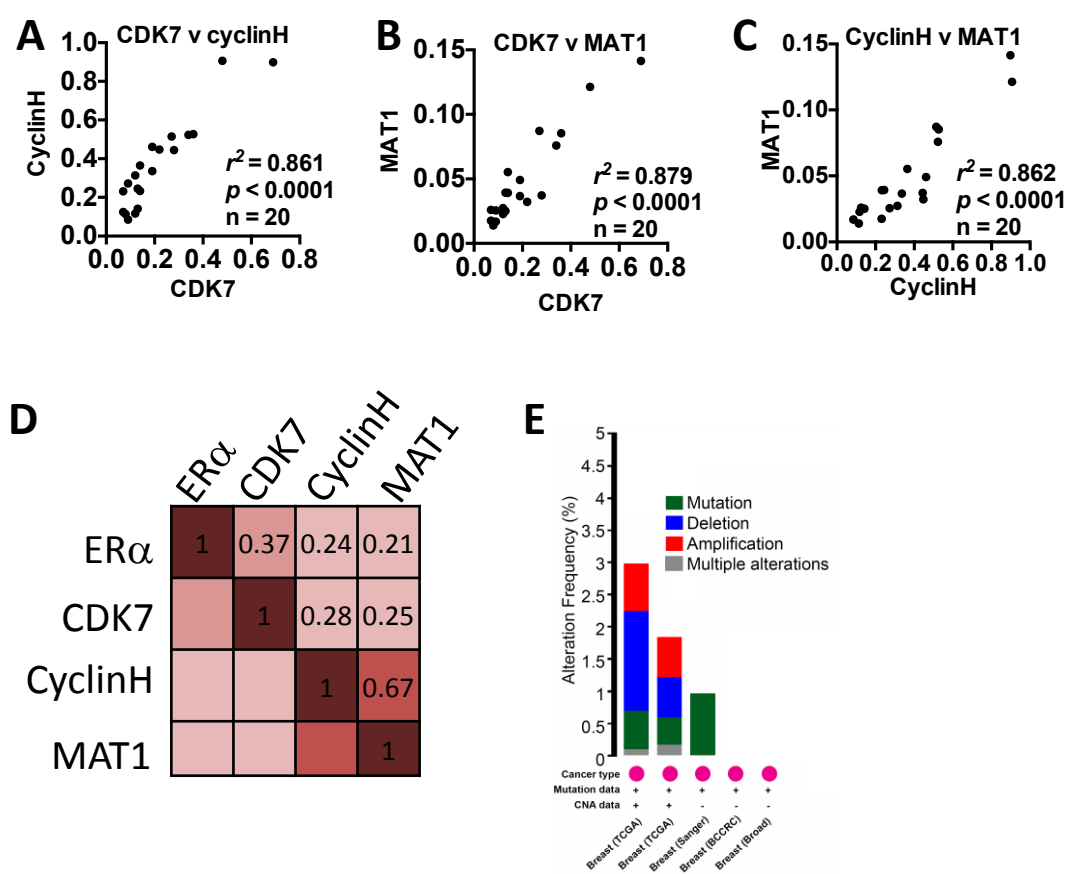
Supplementary Figure 1. Expression of CDK7, cyclin H and MAT1 is elevated in breast cancer.

(A-C) CDK7, CyclinH and MAT1 mRNA levels, determined by real-time RT-PCR analysis, are shown relative to expression of GAPDH for RNA prepared from 20 paired tumour and adjacent normal tissues. Error bars represent SEM for three technical replicates. The paired t-test was used to determine statistical significance of the difference in expression between adjacent normal tissue and paired tumour samples; p-values for differences are indicated. **(D)** Real-time RT-PCR analysis of EPCAM mRNA levels in the 20 paired tumor and adjacent normal samples. Real-time RT-PCR assay (Taqman assay hs00901885; Life Technologies). **(E)** Representative IHC images for tumours that feature adjacent normal elements. **(F)** H-scoring of CDK7, CycH and MAT1 immunostaining in the adjacent normal epithelial cells and cancer cells (n=10).



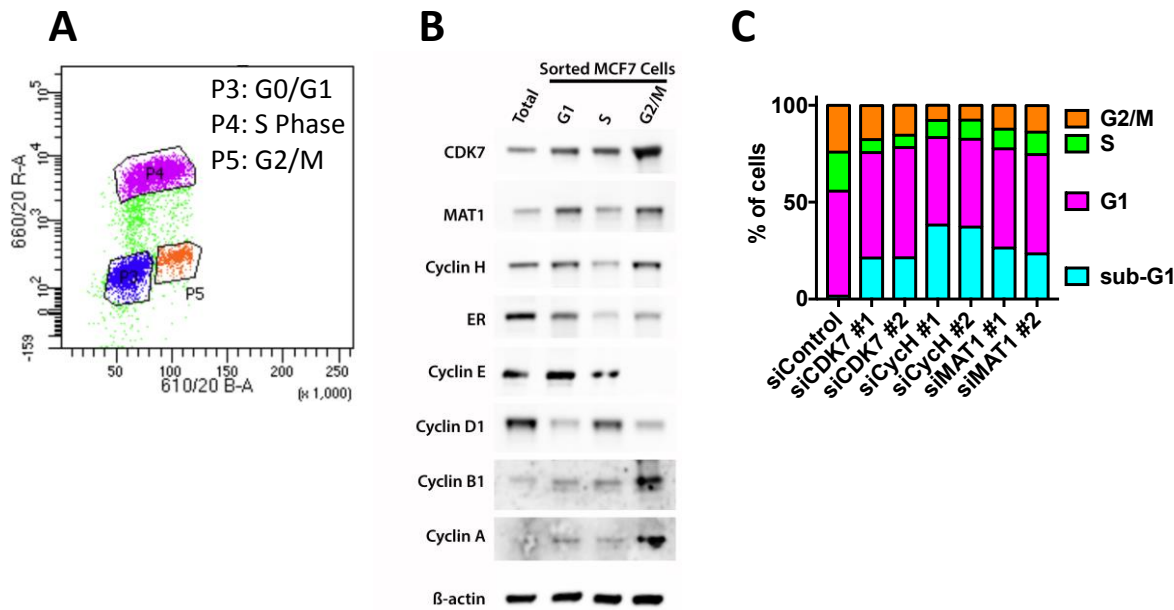
Supplementary Figure 2. Relationship between mRNA levels of CDK7, cyclin H and MAT1 in breast cancer.

(A-C) Plotted are the mean expression values for CDK7 against CyclinH, CDK7 against MAT1 and CyclinH versus MAT1 for each of the 20 tumours samples in Fig. 1A-C. Pearson correlation coefficient r^2 and p-values are shown. **(D)** Pearson's correlation analysis for 1,959 METABRIC breast cancer samples is shown in heat map form, with r^2 values also shown. **(E)** Analysis of TCGA, Sanger, BCCRC and Broad Institute data sets shows that mutations, deletions and/or amplifications in CDK7, CyclinH and MAT1 genes is rare (<3% of cases) in breast cancer.



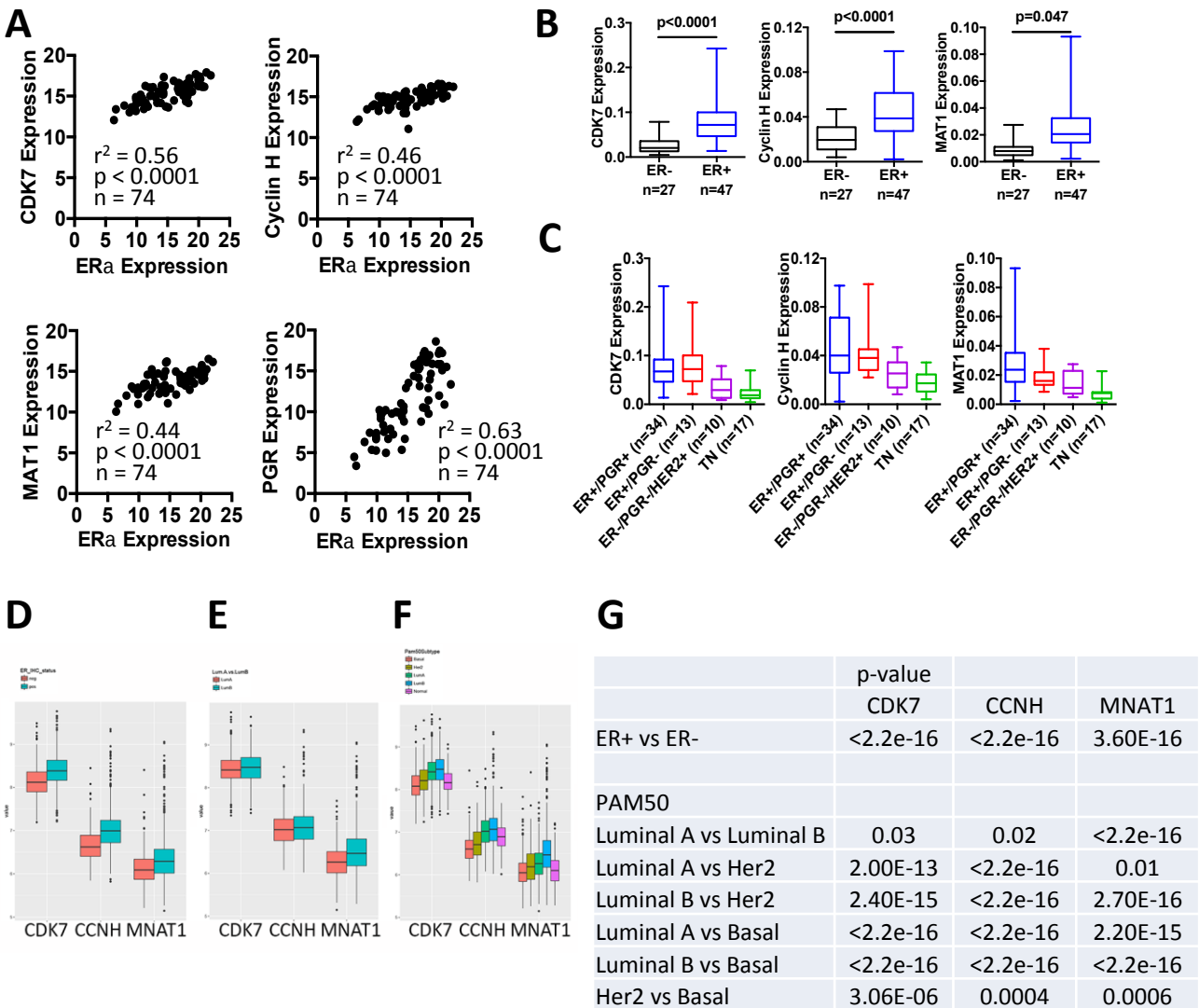
Supplementary Figure 3. Profiling of CDK7, Cyclin H and MAT1 expression in the cell cycle.

- (A) 1×10^6 MCF7 cells in culture treated with 10 μ M EDU (Invitrogen) for 30 mins. were trypsinized, washed with 3ml 1% BSA and fixed with 100 μ l Click-iT™ fixative (component D) for 15 minutes. Fixed cells were washed in 3ml of 1% BSA solution and permeabilised with 100 μ l 1x saponin-based buffer (10x diluted in 1% BSA) for 15 minutes followed by another wash with 3ml of 1% BSA. 500 μ l of reaction cocktail was added to the cells for 30 minutes. Reaction cocktail contained 438 μ l 1xClick-IT® reaction buffer, 10 μ l copper sulphate, 2.5 μ l fluorescent dye azide and 50 μ l reaction buffer additive per sample. Cells were washed in 3ml 1x saponin-based buffer and re-suspend in 500 μ l 1x saponin-based buffer. 5 μ l of RNase A and 2 μ l of PI dye was added to each sample and mixed. Samples were analysed for cell cycle profiles by BD FACS Canto (BD biosciences) and cell cycle phases were sorted using the BD FACS Aria IIu (BD biosciences). Cells gated as marked were collected for protein lysate preparation in RIPA.
- (B) Immunoblotting for cyclin A, cyclin B1, cyclin D1 and cyclin E was used to confirm the phase of the cell cycle of the FACS sorted MCF7 cells. Immunoblotting for CDK7, MAT1 and cyclin H indicates that they are expressed through the cell cycle, with highest levels in G2/M. Cyclin A (sc-751), cyclin B1 (sc-7393) and cyclin D1 (sc-753) antibodies, purchased from Santa Cruz were used at a dilution of 1:250. The Cyclin E antibody (#4129; Cell Signaling) was used at a dilution of 1:500.
- (C) MCF7 cells were treated with siRNAs for CDK7, cyclinH, MAT or with a control siRNA. Cells were used for cell cycle analysis after 48 hours, following fixation and PI labelling.



Supplementary Figure 4. Expression of CDK7, CyclinH and MAT1 is elevated in breast cancer.

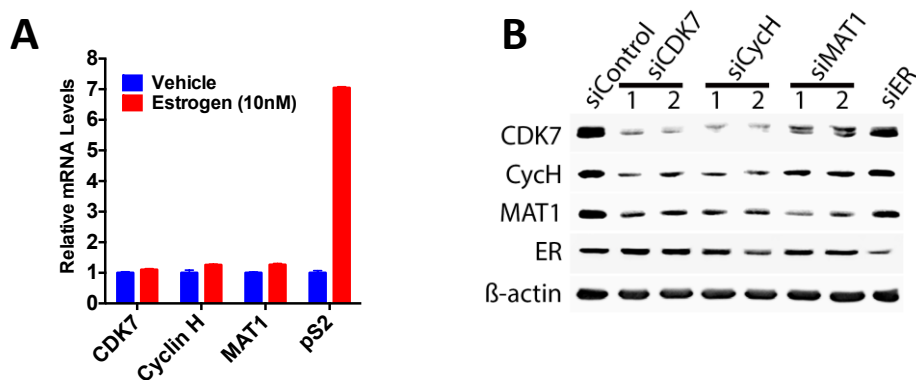
(A) CDK7, CyclinH, MAT1 and ER expression was determined by qRT-PCR in RNA prepared from 74 breast cancer samples. Pearson's correlation values for ER mRNA expression and individual CAK subunit mRNA expression are shown. **(B)** CDK7, cyclinH and MAT1 mRNA levels are plotted relative to GAPDH expression according to clinically defined ER status in the 74 tumour RNAs. **(C)** CAK gene expression is shown according to breast cancer subtypes. Expression of all three subunits was significantly different for ER+/PGR+ vs HER2, ER+/PGR- vs Her2+, ER+/PGR+ vs triple negative (TN), ER+/PGR- vs TN breast cancer ($p < 0.01$). There was no difference in expression of any CAK gene between Her2+ and TN breast cancer, or between ER+/PGR+ and ER+/PGR- samples. **(D-F)** Expression analysis of METABRIC microarray data shows that expression of CDK7, cyclin H (CCNH) and MAT1 (MNAT1) genes is significantly higher in ER-positive ($n=1489$), compared with ER-negative breast cancer ($n=439$) (C). There is higher expression of CAK genes between luminal A ($n=718$) and luminal B ($n=488$) breast cancer (D). Expression of CDK7, cyclin H and MAT1 is significantly higher in luminal A or luminal B breast cancer, compared with basal ($n=329$), HER2 ($n=240$) or 'normal-like' ($n=199$) breast cancer (E). **(G)** Statistical significance was determined using the Mann-Whitney test.



Supplementary Figure 5. CDK7, CyclinH and MAT1 expression is not estrogen regulated in breast cancer cells.

(A) MCF-7 cells cultured in estrogen-free culture medium (DMEM lacking phenol red and supplemented with 10% dextran-coated charcoal stripped fetal calf serum) were treated with estrogen for 16 hours, followed RNA preparation and real-time RT-PCR for CDK7, cyclinH and MAT1, using real-time assays described in Materials and Methods. Real-time RT-PCR assay (Taqman assay hs00170216; Life Technologies) was used for pS2 (TFF1) gene expression. RT-PCR for pS2 was performed as a positive control for an ER target gene (n=3, error bars = SEM).

(B) Immunoblotting of MCF7 cells shows that CDK7, Cyclin H and MAT1 expression is not affected by siRNA mediated ER knockdown. CDK7, cyclin H and MAT1 siRNAs are described in Materials and Methods. siER was purchased from Qiagen (catalogue no. SI03114979).



Supplementary Figure 6. Serine 118 phosphorylation is associated with better prognosis.

(A) Examples of IHC staining for P-Ser118. **(B)** Kaplan-Meier plot showing time to distant metastasis for patients with absent (n=153), weak (n=80), moderate (n=307) or high (n=198) serine 118 phosphorylation. Log rank test (Mantel-Cox) value is shown, together with the p-value. **(C)** Kaplan-Meier plot for breast cancer specific survival in patients with absent (n=149), weak (n=69), moderate (n=253) or high (n=160) serine 118 phosphorylation.

