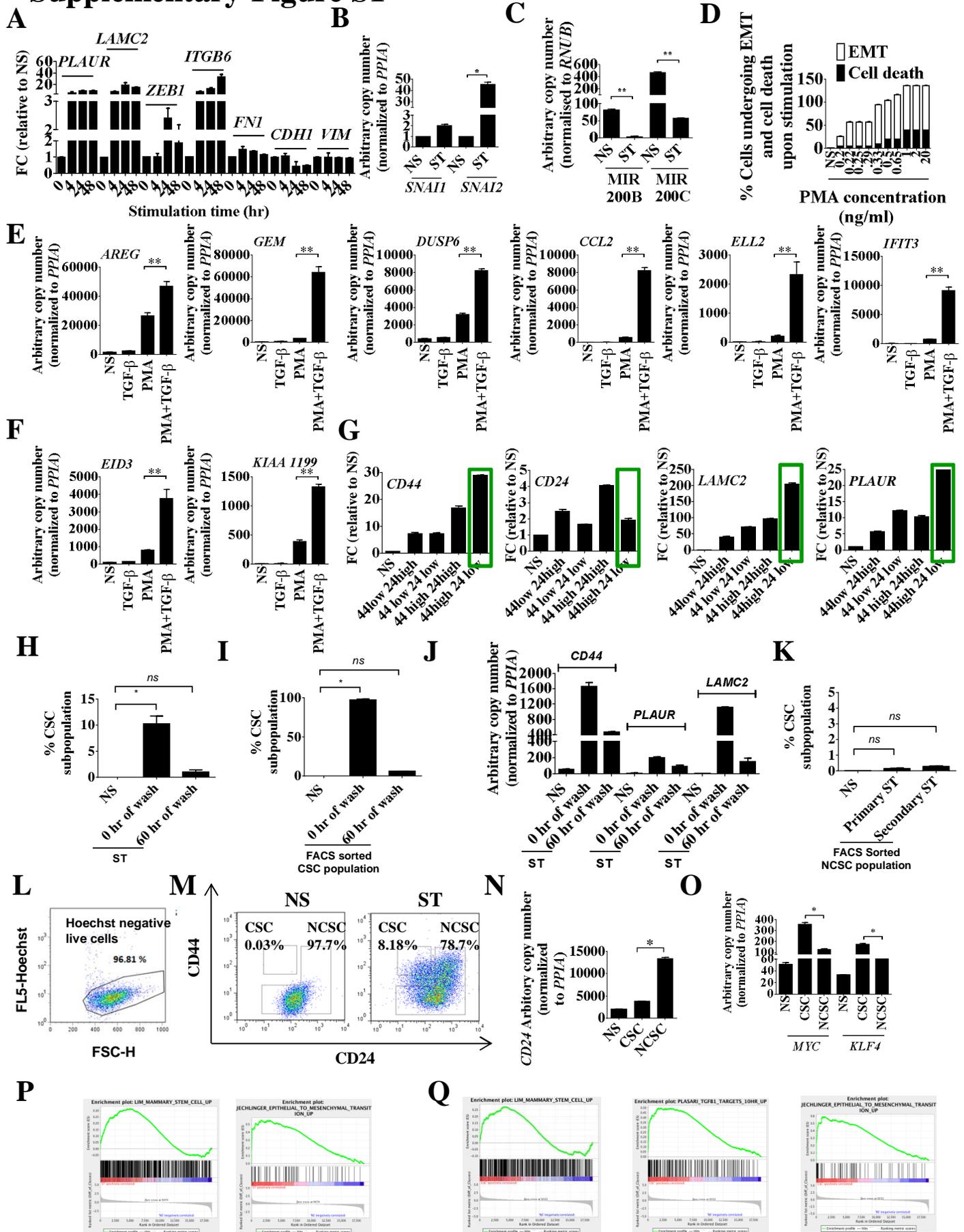
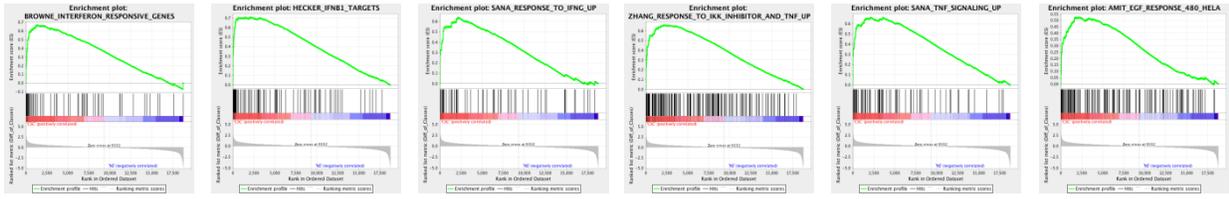


## **SUPPLEMENTARY MATERIALS**

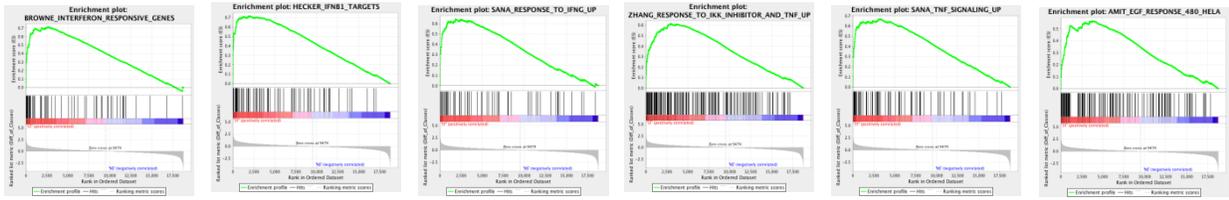
# Supplementary Figure S1



R



S

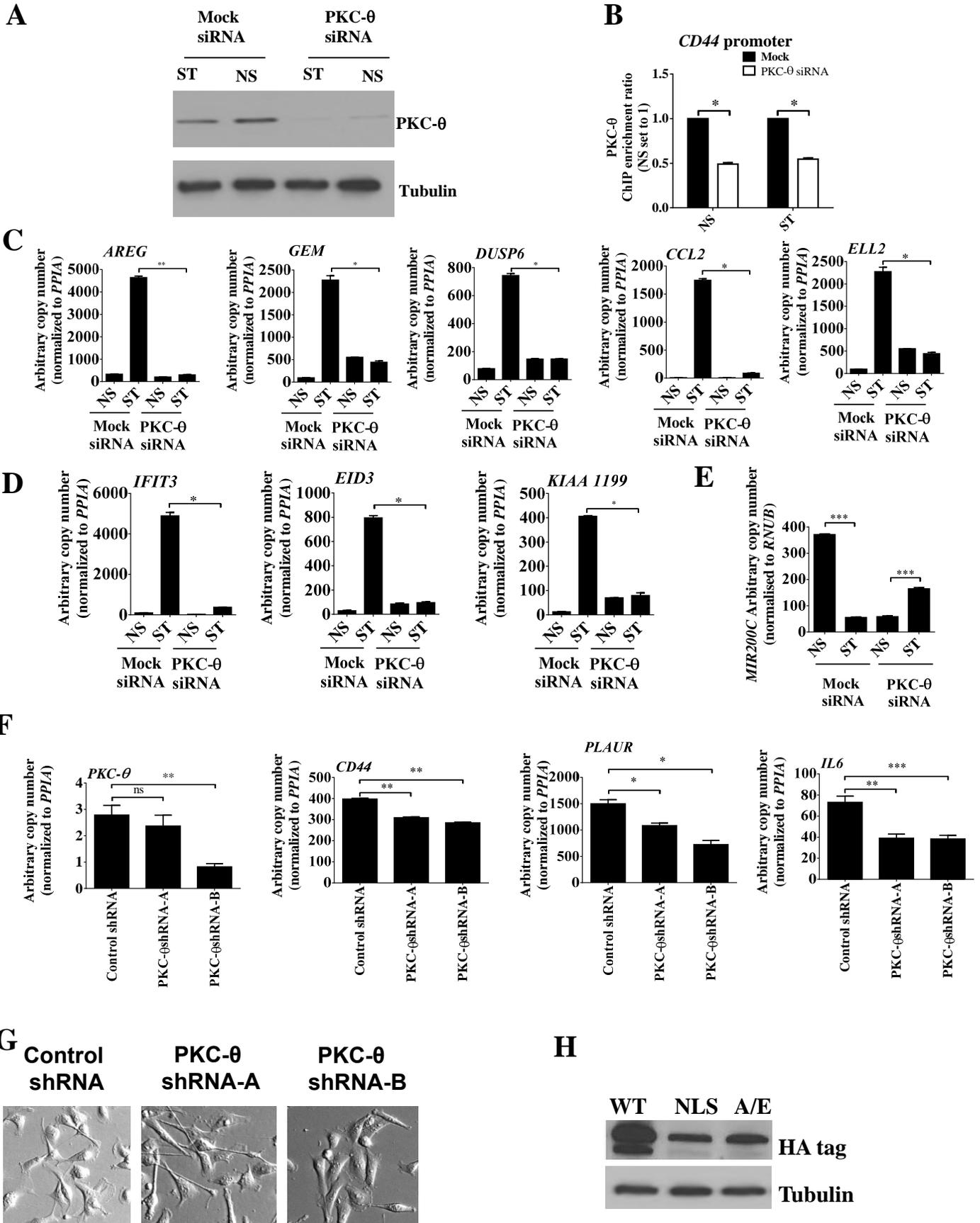


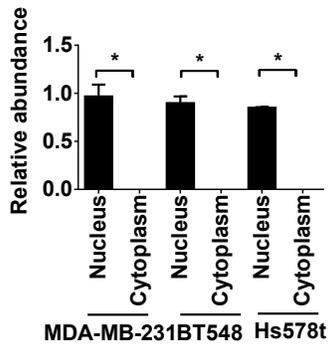
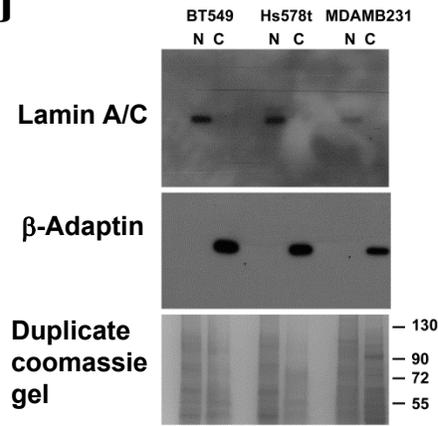
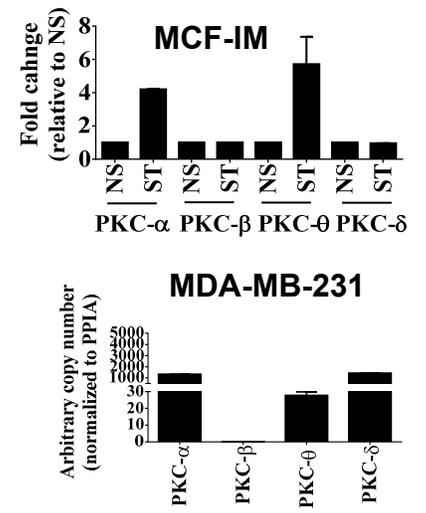
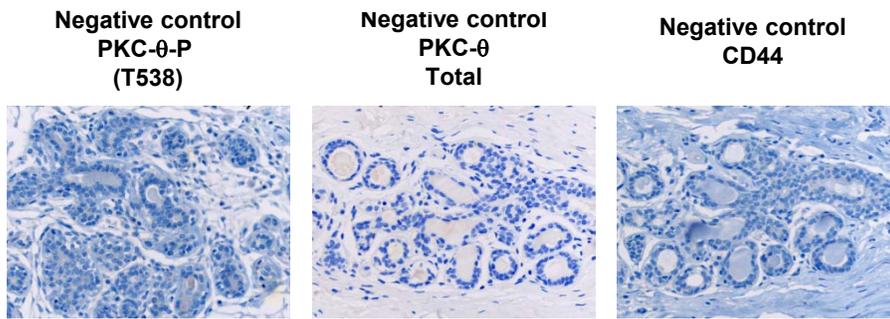
**Figure S1. A CSC population emerges upon PKC pathway induction in the MCF-IM model.**

(A) mRNA kinetics were measured by real-time PCR of cDNA prepared from MCF-7 cells, either non-stimulated (0 h) or after PMA (0.65 ng/ml) stimulation for the times indicated. mRNA levels are expressed as n-fold changes (FC) relative to those of NS samples. (B) mRNA expression of either non-stimulated (NS) or PMA stimulated (ST) MCF-7 cells in MCF-IM for *SNAI1* and *SNAI2*. mRNA levels are expressed as arbitrary copy numbers. (C) Expression of *MIR200* family members of samples in (B) above (D) % Cells undergoing EMT and cell death in MCF-IM. MCF-7 cells were either non-stimulated (NS) or stimulated for 60 h with PMA ranging in concentration between 0.02 ng/ml to 20 ng/ml. For each dilution, three phase contrast images were taken and at least 200 cells were counted in each image before calculating the average of the cells undergoing either percent EMT (as measured by morphology; white bars) or percent cell death (as measured by Vi-CELL-XR cell counter; black bars). (E), (F) mRNA expression of a panel of inducible genes in MCF-IM induced with PMA, TGF- $\beta$  or both. Genes were selected from Affymetrix microarrays detailed Fig. 1C for the MCF-IM. mRNA abundance are expressed as arbitrary copy numbers. (G) mRNA expression of a panel of inducible genes measured by real-time PCR. MCF-7 cells either non-stimulated (NS) or PMA stimulated (0.65 ng/ml for 60 h) and FACS sorted  $CD44^{low}/CD24^{high}$ ,  $CD44^{low}/CD24^{low}$ ,  $CD44^{high}/CD24^{high}$  and  $CD44^{high}/CD24^{low}$  sub-population. (H) %  $CD44^{high}/CD24^{low}$  CSCs subpopulation from MCF-IM model either non-stimulated (NS) or stimulated with PMA (ST) for 60 hr and subsequently washed away PMA at 0 hr or 60 hr post stimulation. (I) %  $CD44^{high}/CD24^{low}$  CSCs subpopulation in FACS sorted CSCs from MCF-IM model. Cells were either non-stimulated (NS) or stimulated (ST) and FACS sorted for CSC at 0 hr of wash or stimulated (ST) and FACS sorted for CSC and CSCs were subsequently left as such for 60 hr. (J) mRNA expression of (F) above for *CD44*, *PLAUR* and *LAMC2*. mRNA abundance are expressed as arbitrary copy numbers. (K) %

CD44<sup>high</sup>/CD24<sup>low</sup> CSCs subpopulation in FACS sorted non-CSCs (NCSCs) from MCF-IM model. Cells were either non-stimulated (NS) or PMA stimulated for the first time (Primary ST) and FACS sorted for NCSC and subsequently left alone or re-stimulated (Secondary ST) and FACS sorted for NCSCs. **(L)** FACS gating strategies to sort Hoechst 33258 negative cells. **(M)** FACS gating strategies to sort CD44<sup>high</sup> /CD24<sup>low</sup> CSCs and NCSCs in MCF-IM cells. % subpopulation is shown above the gates. **(N)** mRNA expression of NS, CD44<sup>high</sup>/CD24<sup>low</sup> CSCs, and NCSCs in MCF-IM for CD24 gene. mRNA abundance are expressed as arbitrary copy numbers. **(O)** mRNA expression of NS, CD44<sup>high</sup>/CD24<sup>low</sup> CSCs, and NCSCs in MCF-IM for a panel of stemness related genes. mRNA abundance are expressed as arbitrary copy numbers. **(P)** Enrichment of the sets of genes up-regulated in mammary cancer stem cells, up-regulated by TGF- $\beta$  in embryonic fibroblasts or epithelial to mesenchymal transition in the list of genes ranked from higher (red) to lower (blue) in CSCs compared to NS in MCF-IM. **(Q)** Enrichment of the sets of genes up-regulated in mammary cancer stem cells or epithelial to mesenchymal transition in the list of genes ranked from higher (red) to lower (blue) in stimulated MCF-IM. **(R)** Enrichment of the sets of genes up-regulated in interferon response, TNF or EGF in the list of genes ranked from higher (red) to lower (blue) in CSCs compared to NS in stimulated MCF-IM. **(S)** Enrichment of the sets of genes up-regulated in interferon response, TNF or EGF in the list of genes ranked from higher (red) to lower (blue) in CSCs compared to NS in MCF-IM. All results represent either the mean  $\pm$  the standard error of three independent experiments or a representative experiment from three replicates (N=3) (J, L). \*\*, P<0.01, \*, P<0.05 and ns, not significant.

# Supplementary Figure S2



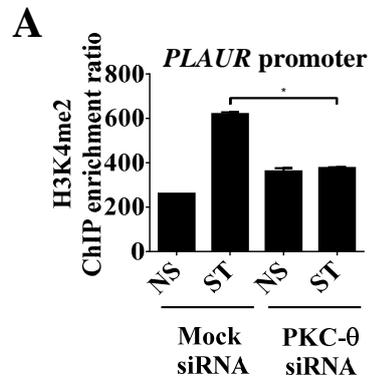
**I****J****K****L**

**Figure S2. Validation of PKC- $\theta$  knockdown by PKC- $\theta$  siRNA and PKC- $\theta$  shRNA.**

(A) Immunoblotting for PKC- $\theta$  after PKC- $\theta$  siRNA treatment in MCF-IM. (B) ChIPs of MCF-IM following transfection with either mock siRNA or PKC- $\theta$  siRNA and subsequently either non-stimulated (NS) or PMA stimulated (ST) with anti-PKC- $\theta$  antibody across *CD44* promoter. (C) and (D) mRNA expression of a panel of inducible genes in MCF-IM following transfection with either mock siRNA or PKC- $\theta$  siRNA and subsequently either left alone (NS) or PMA stimulated (ST). Genes were selected from Affymetrix microarrays detailed Fig. 2A in the MCF-IM model. mRNA abundance are expressed as arbitrary copy numbers. (E) *MIR200C* expression in samples as (A) above. (F) mRNA expression of PKC- $\theta$  and a panel of inducible CSC enriched-EMT genes in the MDA MB-231 cells after transfection with either control shRNA or PKC- $\theta$  shRNA-A or PKC- $\theta$  shRNA-B. mRNA abundance are expressed as arbitrary copy numbers. (G) Representative phase contrast images of (F). (H) Immunoblotting for HA tagged PKC- $\theta$  wild type (WT), nuclear localization sequence (NLS) and kinase dead mutant with intact nuclear localization signal (A/E) constructs. Tubulin was used as a loading control. (I) Abundance of the PKC- $\theta$  protein in nuclear and cytoplasmic lysates of the indicated cell lines as measured by Western blot. Abundances are relative to the loading control  $\beta$ -Adaptin. (J) Immunoblotting controls of nuclear (Lamin A/C), cytoplasmic ( $\beta$ -adaptin) loading controls and duplicate coomassie gel for indicated metastatic cell lines. (K) mRNA expression of PKC isozymes measured by real-time PCR in MCF-IM and MDA-MB-231 cells. MCF-7 cells were either non-stimulated (NS) or PMA stimulated (0.65 ng/ml) for 60 hours. mRNA levels are expressed as n-fold changes (FC) relative to those of NS samples. (L) Photomicrographs of antibodies used in the immunohistochemistry studies. A section of breast cancer and normal breast tissue was run as a negative control by omitting primary antibody. All results represent either the mean  $\pm$  the standard error of three

independent experiments or a representative experiment from three replicates (N=3) (F). \*\*, P<0.01, \*, P<0.05 and ns, not significant.

# Supplementary Figure S3

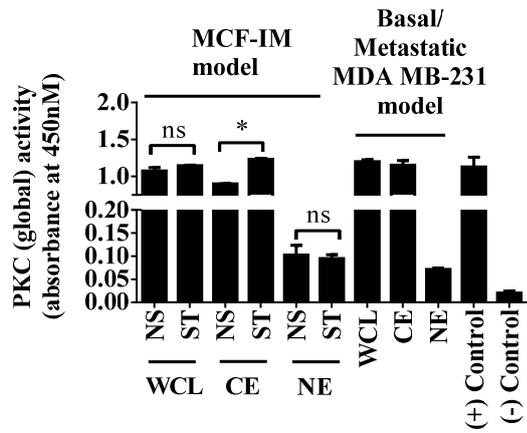


**Figure S3. Knockdown of PKC- $\theta$  decreased chromatin association of active chromatin mark.**

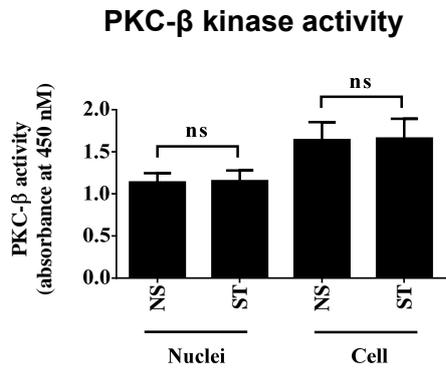
(A) ChIP as in Fig. 1H with anti-H3K4Me2 antibody after PKC- $\theta$  siRNA treatment in MCF-IM cells. Results represent the mean  $\pm$  the standard error of three independent experiments (N=3). \*\*, P<0.01, \*, P<0.05 and ns, not significant.

# Supplementary Figure S4

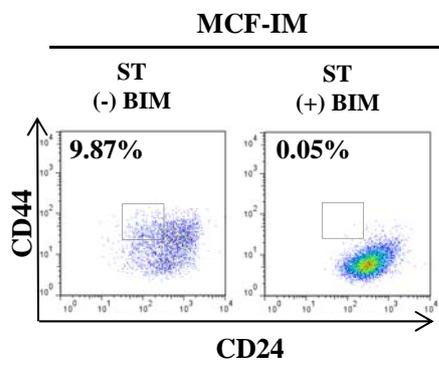
**A**



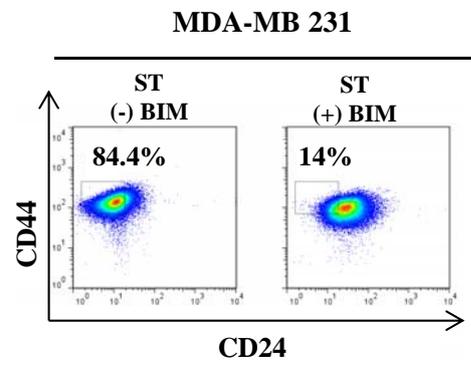
**B**



**C**

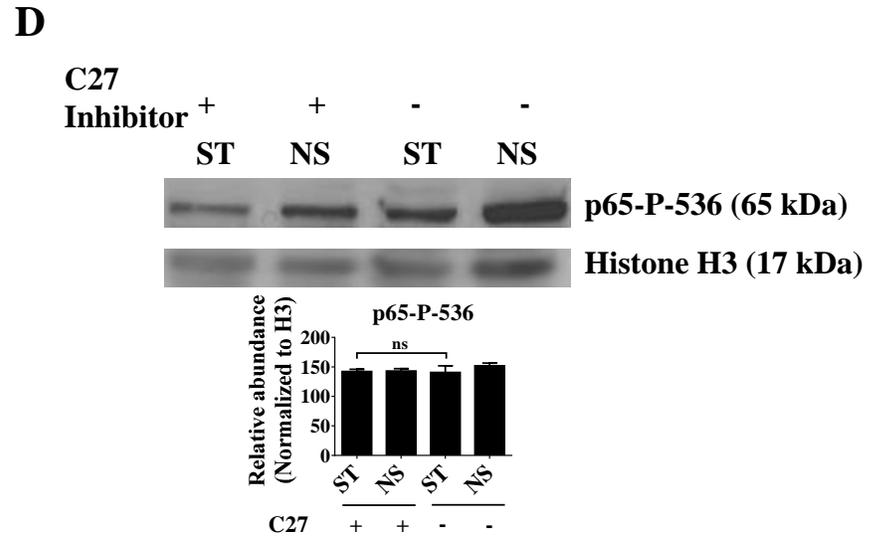
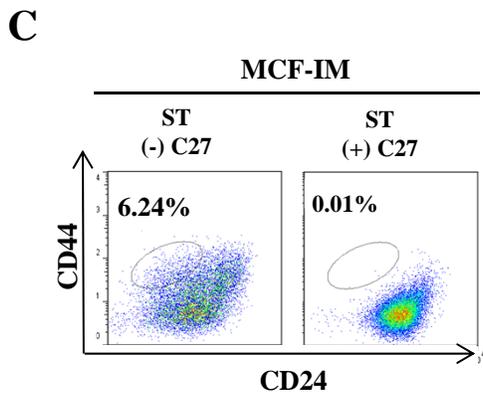
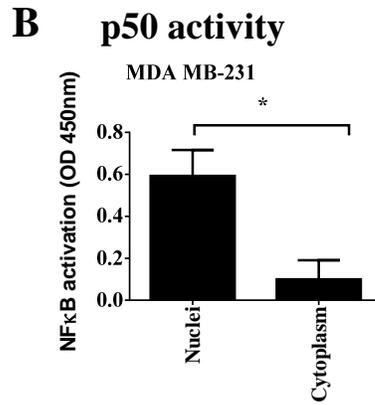
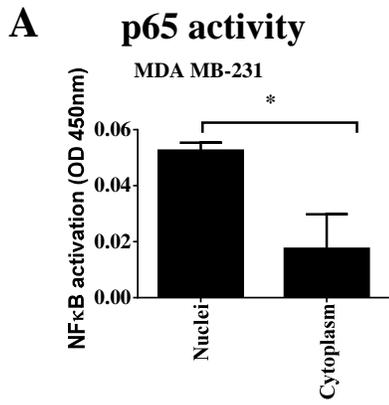


**D**



**Figure S4. PKC-θ activity inhibition abolishes CD44<sup>high</sup>/CD24<sup>low</sup> CSC formation in MCF-7 and MDA-MB-231 cell lines.**

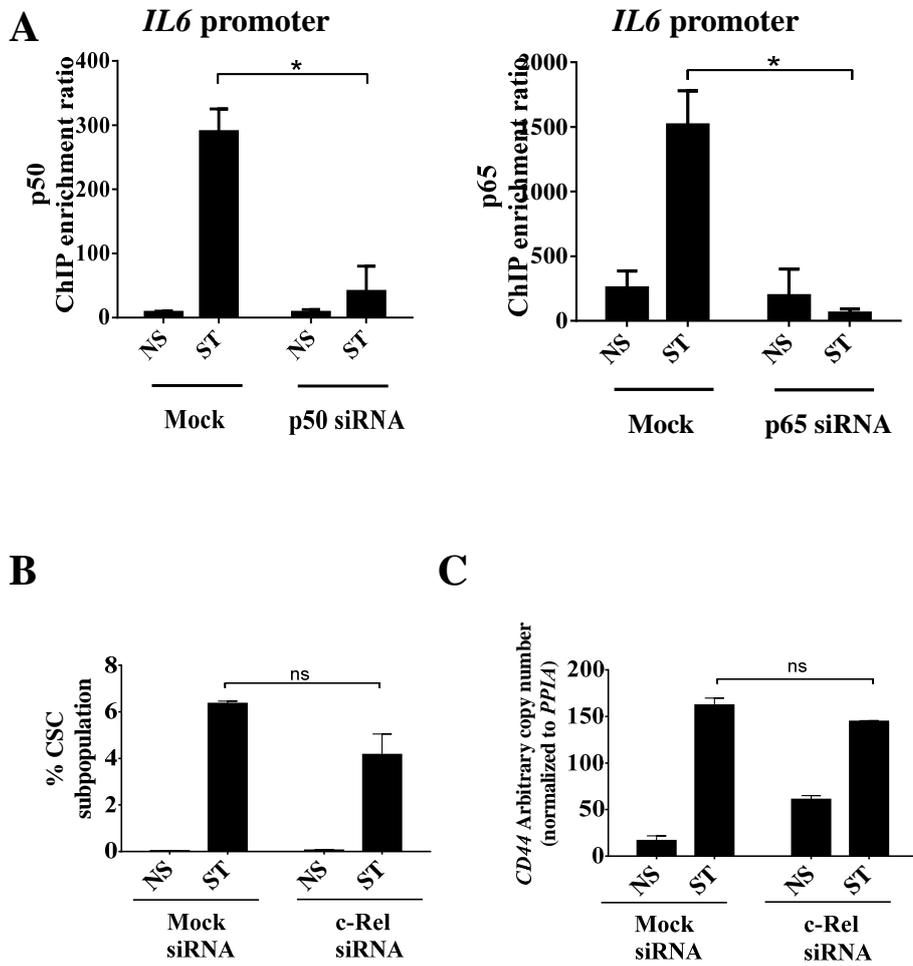
(A) PKC ELISA-based kinase assays of whole cell lysate (WCL), cytoplasmic extract (CE), or nuclear extract (NE) from MCF-IM and MDA-MB-231 cells. Assays were carried without specific immunoprecipitates for global PKC activity. (B) PKC ELISA-based kinase assays of nuclei and whole cell of MCF-IM using anti-PKC-β antibody for immuniprecipitation. Kinase assays were performed on these PKC-β fractions and absorbance was measured at 450 nM. Data represented as kinase activities relative to negative control. Recombinant active PKC (positive (+) control); secondary antibody alone (negative (-) control). (C) and (D) Representative FACS plot showing CD44<sup>high</sup>/CD24<sup>low</sup> CSC subpopulation in MCF-IM model and MDA-MB-231 cells respectively. MCF-7 cells were stimulated with PMA (0.65 ng/ml for 60 h) either without pre-treatment of PKC activity inhibitor bisindolylmaleimide-I (**ST-BIM**) or with pre-treatment with bisindolylmaleimide (**ST+BIM**). Representative experiment from three replicates (N=3) has been shown.



**Figure S5. PKC- $\theta$  regulates p65 and p50 activity in MCF-IM**

(A) p65 activity and (B) p50 activity of MDA-MB-231 cells using ELISA based NF- $\kappa$ B activity assays. (C) Representative FACS plot showing CD44<sup>high</sup>/CD24<sup>low</sup> CSC subpopulation in MCF-IM model. MCF-7 cells were stimulated with PMA (0.65 ng/ml for 60 h) either without pre-treatment of PKC- $\theta$  specific activity inhibitor compound 27 (-C27) or with pre-treatment with compound 27 (+C27) (1 $\mu$ M for 24 hours). (D) Immunoblotting for of MCF-IM nuclear extracts phosphorylated NF- $\kappa$ B p65 at serine 536 (p536) after with (+) or without (-) compound 27 treatment. Densitometric analyses of (D) using Image J software provided below western blot. 5 $\mu$ g of the nuclear extract was used for western blots and histone H3 antibody was used as a nuclear control. Representative experiment from three replicates (N=3) has been shown.

# Supplementary Figure S6

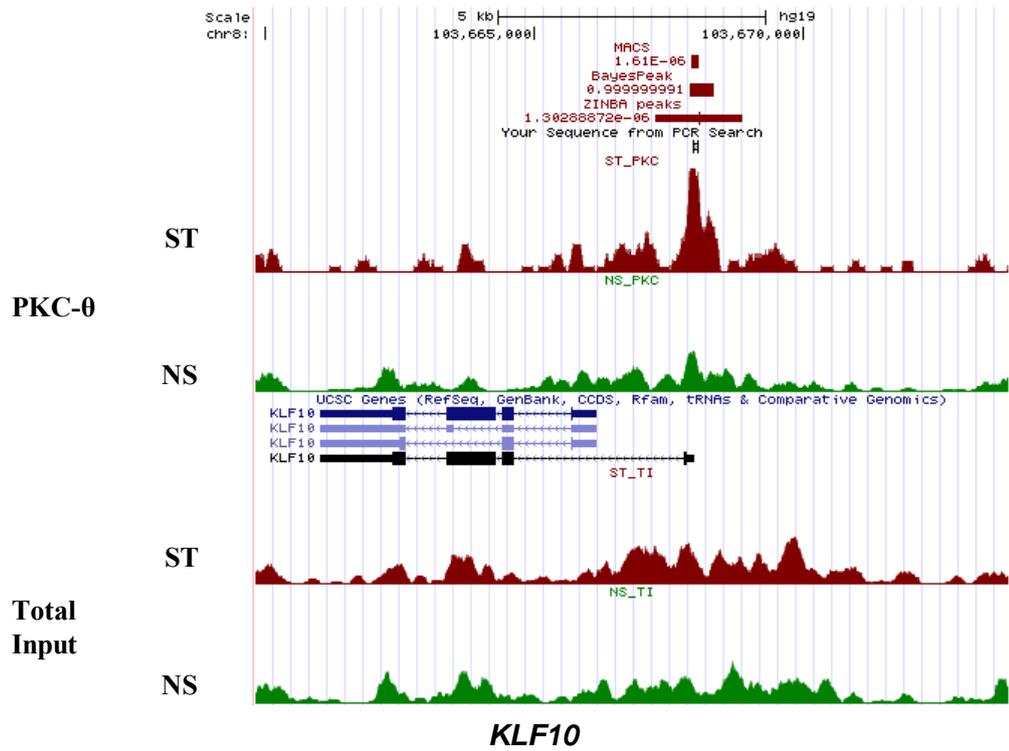


**Figure S6. Validation of p50 and p65 protein knockdown by siRNA.**

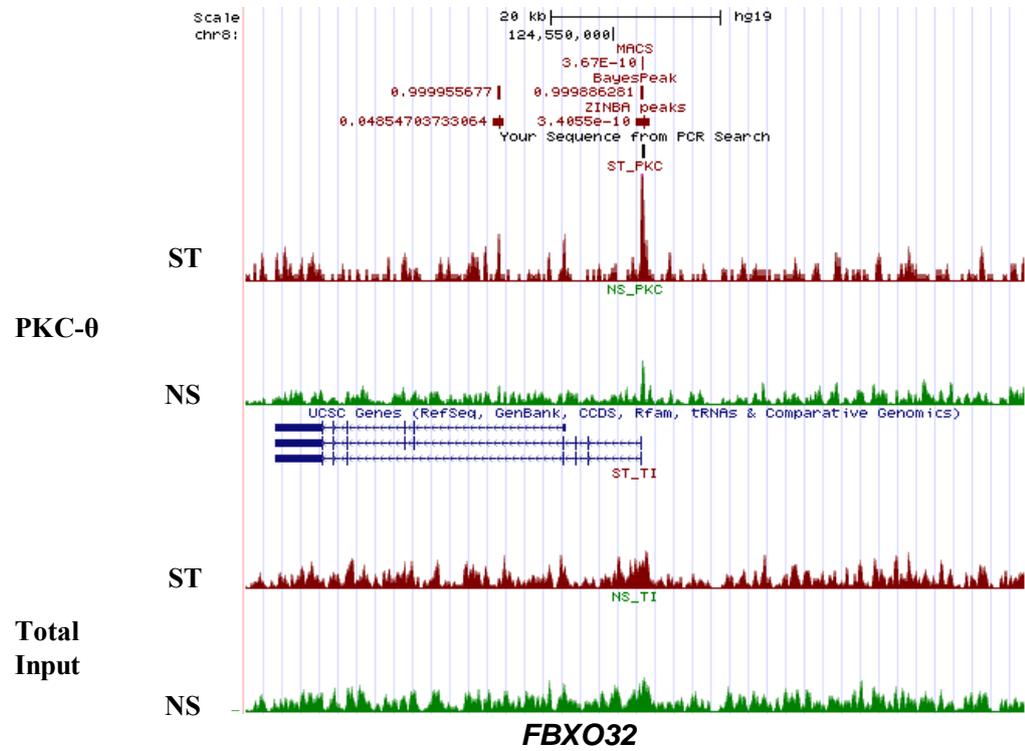
(A) p50 or p65 ChIPs in MCF-IM across *IL6* promoter regions following transfection with either mock siRNA or p50 siRNA or p65 siRNA and subsequently either non-stimulated (NS) or PMA stimulated (ST). (B) CD44<sup>high</sup>/CD24<sup>low</sup> CSCs FACS analysis of MCF-IM transfected with either mock siRNA or c-Rel siRNA and subsequently either left alone (NS) or stimulated (ST) for 60 hr. (C) mRNA expression by real-time PCR of *CD44* of samples in (B) above. Data represent the mean of fold change  $\pm$  standard error (SE) of three independent experiments. All results represent the mean  $\pm$  the standard error of three independent experiments (N=3). \*\*, P<0.01, \*, P<0.05 and ns, not significant.

# Supplementary Figure S7

A

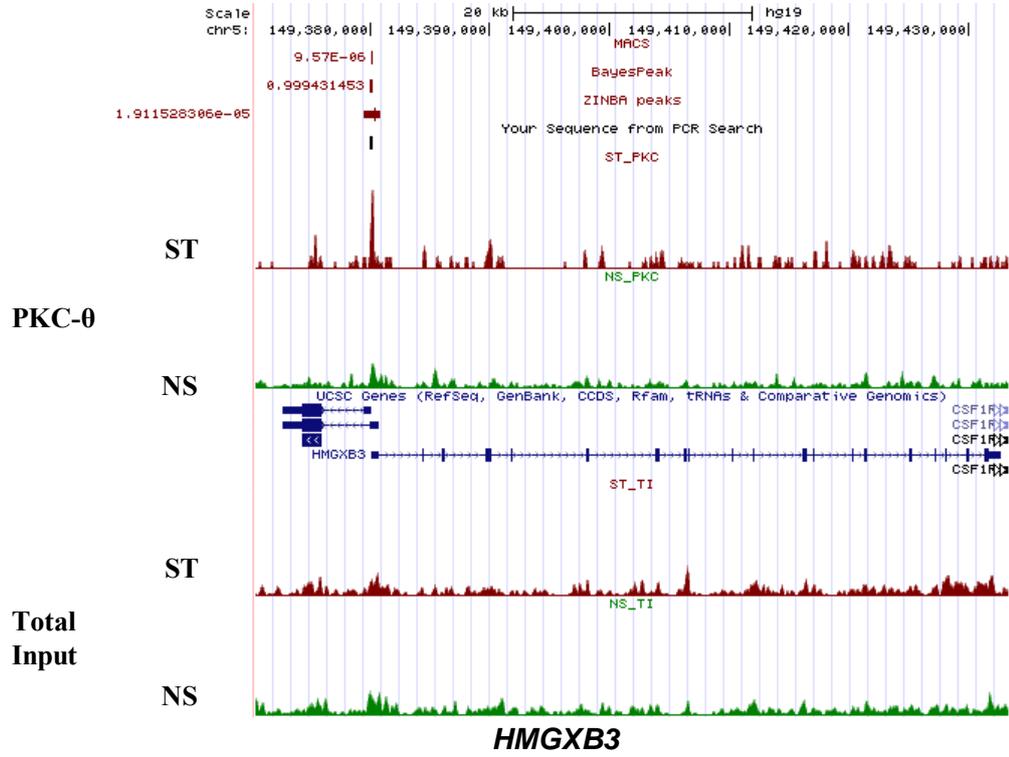


B

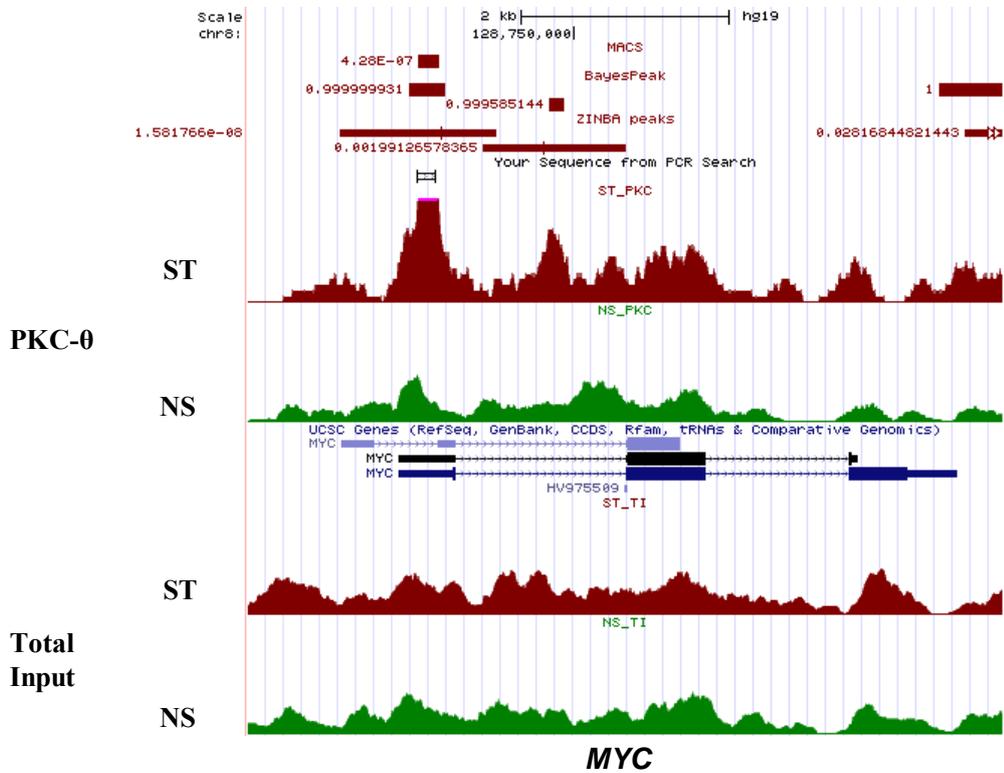


# Supplementary Figure S7

C

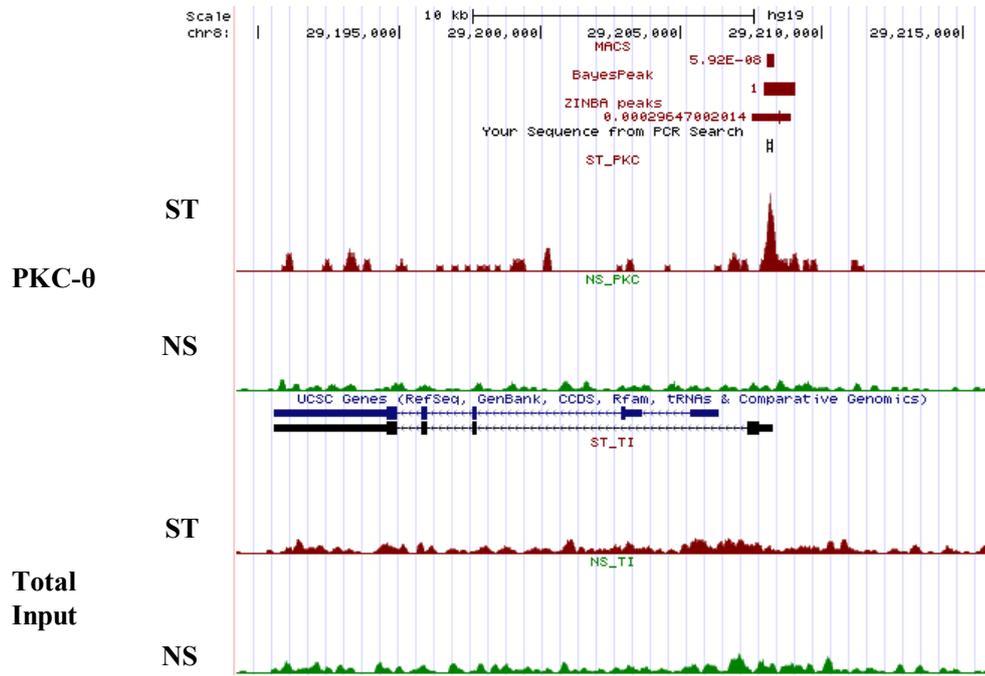


D



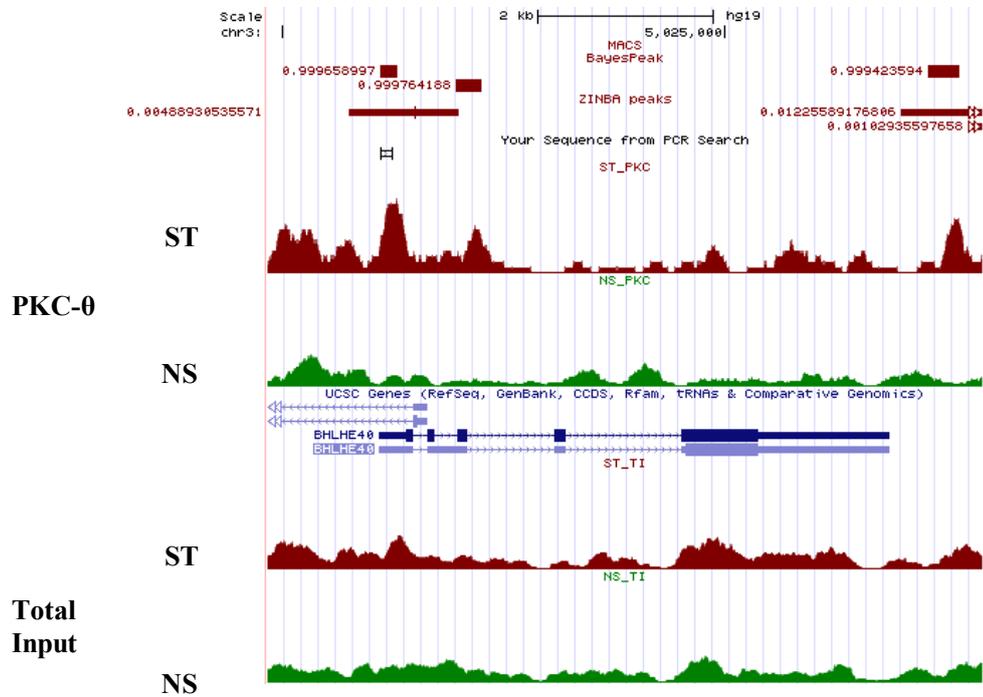
# Supplementary Figure S7

**E**



**DUSP4**

**F**



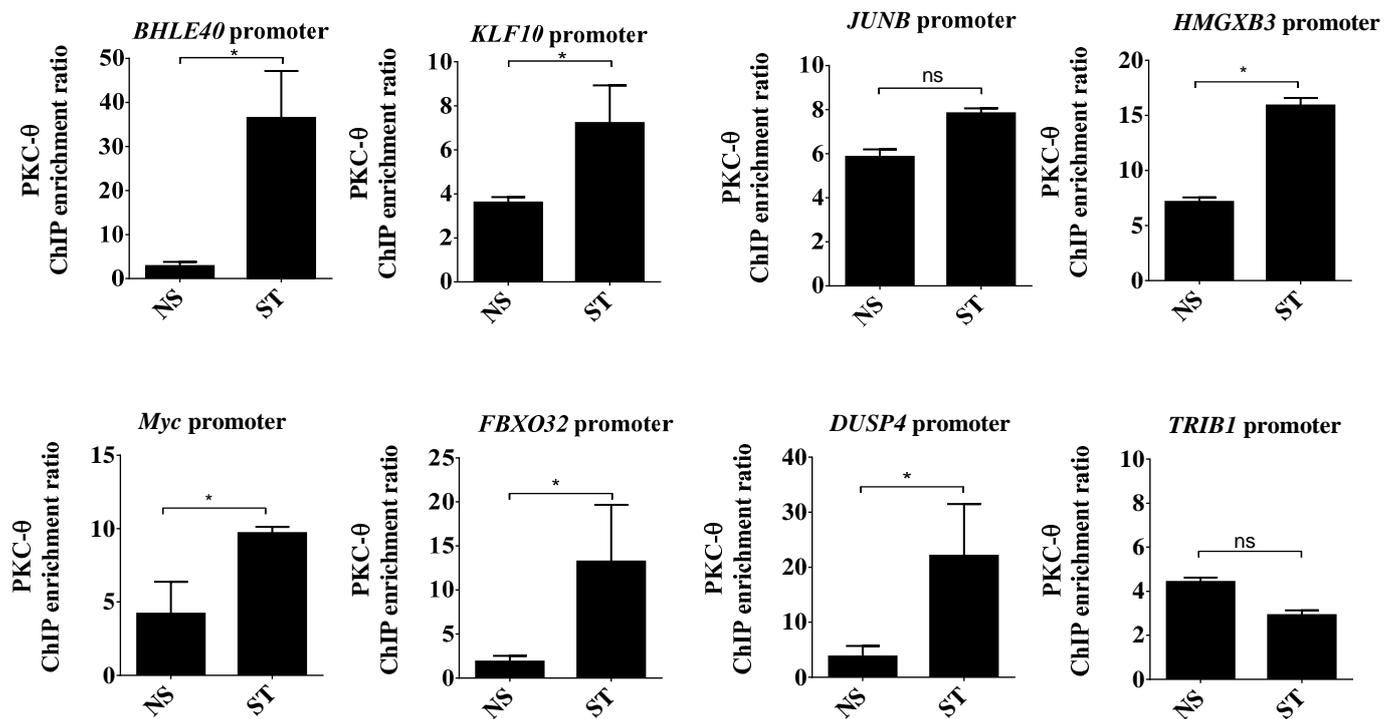
**BHLHE40**

**Figure S7. ChIP-seq as compared by three peak caller programs, BayesPeaks, ZINBA and MACS2.**

PKC- $\theta$  binding at (A) *KLF10*, (B) *FBXO32*, (C) *HMGXB3*, (D) *MYC*, (E) *DUSP4* and (F) *BHLHE40* in non-stimulated (NS) and PMA stimulated (ST) MCF-7 cells. Peaks called by MACS2 and ZINBA are shown with their p-values while peaks called by BayesPeak are shown with posterior probabilities.

# Supplementary Figure S8

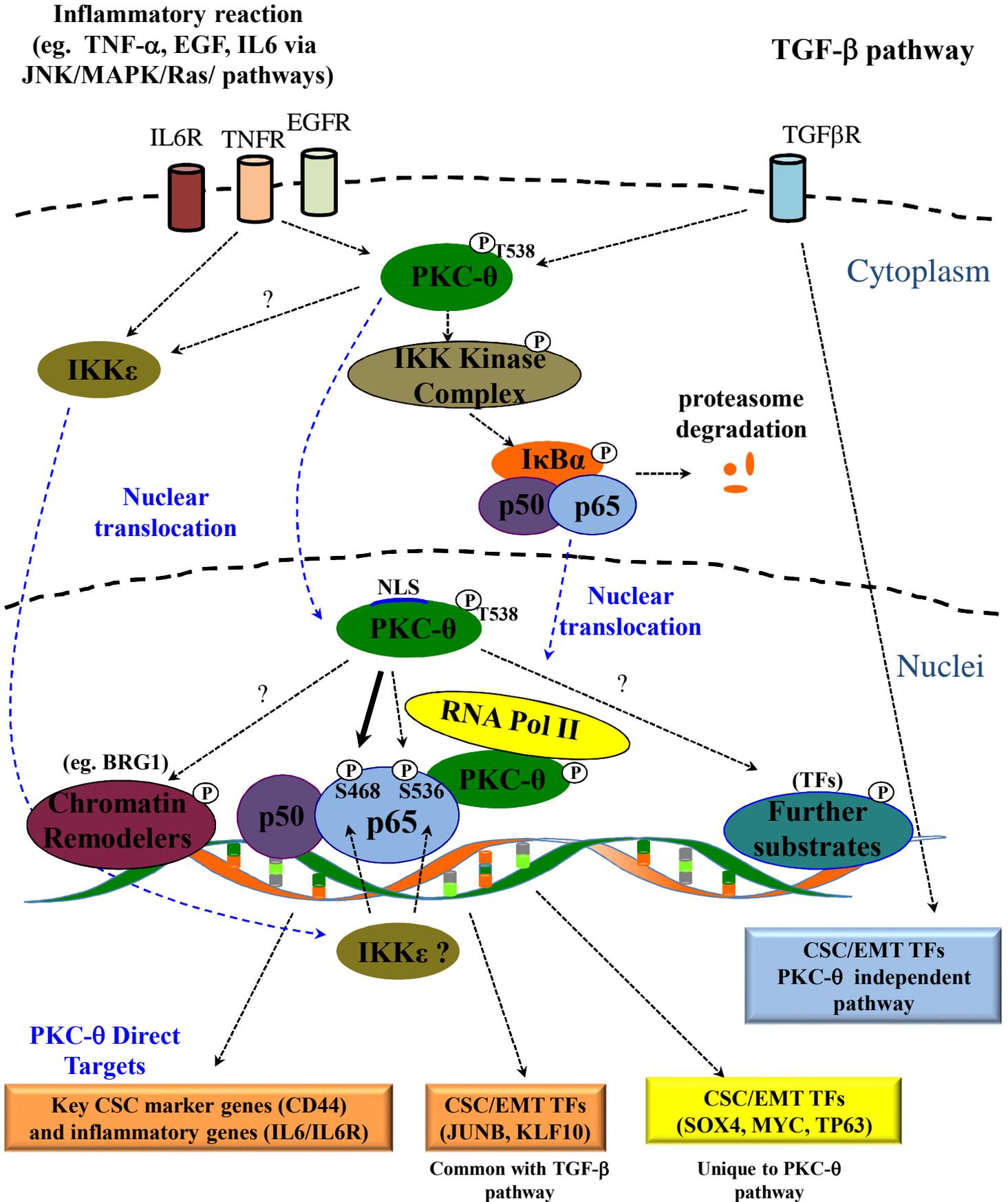
A



**Figure S8. Validation of PKC- $\theta$  ChIP sequencing analysis by real time-PCR.**

(A) PKC- $\theta$  ChIPs on MCF-7 cells either non-stimulated (NS) or stimulated with PMA (ST) for 60 hr across a panel of PKC- $\theta$  bound inducible genes selected from ChIP sequencing analysis from Figure 7D. All results represent either the mean  $\pm$  the standard error of three independent experiments (N=3).

## Mesenchymal / CSC state



**Figure S9: A model to show the dual roles for PKC- $\theta$  signaling and chromatin regulation in CSC/EMT inducible gene regulation via inflammatory transcription factors (TFs).**

We propose a model whereby PKC- $\theta$  integrates inflammatory and TGF- $\beta$  signals to control the transcription of a distinct cohort of inducible genes: EMT/CSC marker genes (e.g., *CD44*, *UPAR*); inflammatory cytokine genes (e.g., *IL6*, *ILR*); master EMT/CSC transcription regulators (e.g., *MYC*, *KLF10*, *JUNB*). PKC- $\theta$  plays a dual role in inducible gene regulation in the mesenchymal/CSC state:

(i) Indirect ‘signaling’ role: PKC- $\theta$  is activated by inflammatory mediators such as TNF $\alpha$ , EGF, and/or TGF- $\beta$  pathways, leading to the activation and nuclear translocation of the central inflammatory driver, NF- $\kappa$ B; this is akin to the role of PKC- $\theta$  kinase signaling in T cells (refs).

(ii) Direct ‘chromatin-tethered’ role: An active p50:p65 NF- $\kappa$ B heterodimer complex, triggered by the PKC- $\theta$  signaling cascade, docks at PKC- $\theta$ -sensitive inducible gene-specific loci and recruits active PKC- $\theta$ . At these loci, PKC- $\theta$  is required for the phosphorylation of p65 at ser-468. Additionally, it is proposed that PKC- $\theta$  phosphorylates other TFs that are recruited to these gene loci, further stabilizing their association with chromatin. It remains to be determined whether chromatin-tethered IKK- $\alpha$ , a nuclear kinase shown to phosphorylate p65 at ser-468 via TNF $\alpha$  and overexpressed in breast cancers, regulates a common or distinct gene subset to that of PKC- $\theta$ . Chromatin-anchored PKC- $\theta$  is essential for the recruitment of the RNA Pol II active transcriptional complex at such loci. This novel ‘PKC- $\theta$ -anchored chromatin complex’ is likely to consist of key CSC transcriptional drivers, such as STAT3, MYC, ZEB, and also chromatin remodeling machinery such as BRG1 and the acetyltransferase GCN5. In the future, proteomic-based strategies will allow us to establish

the precise components of the PKC- $\theta$  chromatin-tethered complex. PKC- $\theta$  is also required for chromatin accessibility at such loci in the mesenchymal state, and we suggest that this occurs either via the recruitment of chromatin remodelers, or by direct phosphorylation of TFs or histones across neighboring nucleosomes.

**Table S1. Gene listing of MCF-IM inducible gene expression microarray profile.**

Microarray analysis was performed on non-stimulated and PMA stimulated cells for 60 h.

Supplementary Table directly relates to Fig. 1C. (File name: Supplementary Table

S1\_Zafar.xlsx).

**Table S2. GSEA report for stimulated (ST) vs non-stimulated (NS).**

Gene Set Enrichment Analysis (GSEA) with sets of genes from the chemical and genetic permutations group MSigDB, on the genes ranked by difference (Log values) between PMA stimulated (ST) and non-stimulated (NS) cells with p-value < 0.01 and FDR < 25%. (File name: Supplementary Table S2\_Zafar.xlsx).

**Table S3. GSEA report for cancer stem cells (CSC) vs non-stimulated cells (NS).**

Gene Set Enrichment Analysis (GSEA) with sets of genes from the chemical and genetic permutations group MSigDB, on the genes ranked by difference (Log values) between cancer stem cells (CSC) and non-stimulated (NS) cells with p-value < 0.01 and FDR < 25%. (File name: Supplementary Table S3\_Zafar.xlsx).

**Table S4. Gene listing of PKC-driven inducible genes that are sensitive to the PKC- $\theta$  pathway.** Microarray analysis was performed on mock and PKC- $\theta$  siRNA-treated MCF-IM model. Supplementary Table directly relates to Fig. 2A. (File name: Supplementary Table S4\_Zafar.xlsx).

**Table S5. Gene listing comparing PKC- $\theta$  binding regions by ChIP sequencing with gene expression microarray analysis from HMLE-IM and MCF-IM.** Supplementary Table directly relates to Fig. 7 A-D. (File name: Supplementary Table S5\_Zafar.xlsx).

**Table S6. Identification of over-represented transcription factor motifs within PKC- $\theta$  binding genes.** The size of the PKC- $\theta$  binding regions for the DPDs was adjusted to 1kb

before being analyzed for over-represented transcription factor binding sites using Genomatix. Twenty-eight motif families had z scores  $> 5$  when their frequency was compared to both promoter regions (which have similar GC content) and background genomic regions. Supplementary Table directly relates to Fig. 7D. (File name: Supplementary Table S6\_Zafar.xls).