An inflammatory regulatory network mediated by the joint action of NF-kB, STAT3, and AP-1 factors is involved in many human cancers

Zhe Ji[1,2,4$], Lizhi He[1$], Aviv Regev[2,3] and Kevin Struhl[1]

SI Appendix

Materials and Methods

Cell culture. The inducible model of cellular transformation involves MCF-10A, a non-transformed mammary epithelial cell line (1) containing ER-Src, a derivative of the Src kinase oncoprotein (v-Src) that is fused to the ligand-binding domain of the estrogen receptor (2). Cells were cultured in DMEM/F12 medium with the supplements as previously described (3, 4). Tamoxifen (Sigma, H7904, 0.4 mM) was used to transform this inducible cell line, when the cells were grown to 30% confluence. For the drug inhibition experiments, cells were pretreated with SP600125 (JNK inhibitor, Selleck Chemical, S1460, 0.3 μM), S-Ruxolitinib (JAK inhibitor, Cayman Chemical, #11609, 0.1 μM), Tofacitinib (JAK inhibitor, Selleck Chemical, S2789, 0.1 μM) and IL-1RA (IL-1R antagonist, Peprotech, #200-01RA, 30 ng/ml) for 75 minutes after which cells were transformed by addition of 0.4 mM Tamoxifen and 4 ng AZD0530 for 24 hours. Transformation efficiency was measured by cell growth under low attachment conditions as described previously (5).

CRISPR Knockouts. CRISPR-blasticidin lentiviral plasmid was constructed by replacing puromycin resistance gene with blasticidin resistance gene in LentiCRISPR V2 plasmid (Addgene, #52961). The oligonucleotide sequences used to clone into CRISPR-blasticidin plasmid were as follows: STAT3 (CACCGCGATCTAGGCAGATGTTGGG and AAACCCCAACATCTGCCTAGATCGC); JUN (CACCGGCACCTCCGCGCCAAGAACT and AAACAGTTCTTGGCGCGGAGGTGCC); JUNB (CACCGGCGCTTTGAGACTCCGGTAG and AAACCTACCGGAAGTCTCAAAGCGCC); NF-κB (CACCGGAATGACAGAGGCGTGTATA and AAACTATACACGCCTCTGTCATTCC); FOS (CACCGGGCGTTGTGAAGACCATGAC and AAACGTCATGGTCTTCACAACGCCC). CRISPR-blasticidin plasmid and three lentiviral plasmids, VSV-G, GP and REV were co-transfected into 293T cells to produce lentiviruses as the previous publication (6). After CRISPR lentiviruses infection, ER-Src cells were selected with blasticidin 10 μg/ml for 3 days to generate CRISPR knockout stable cell lines. The knockout efficiencies of the transcription factors were assessed by Western blotting.

Nucleocytoplasmic separation, co-immunoprecipitation, and Western blotting. Cells were suspended in a buffer (10 mM HEPES PH 7.5, 10 mM KCl, 2 mM MgCl, PMSF 0.1 mM and Roche complete protease inhibitor, Sigma, #11697498001) and incubated on ice for 20 min. After grinding the cells 50 times with a Wheaton Dounce A, the cell lysate was layered on top of 40% sucrose buffer and centrifuged at top speed at 4^0C for 5 minutes. The supernatant (cytoplasm) and pellet (nucleus) were separated, and the The nuclear pellet was re-suspended in 20 mM Tris pH 7.4, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 25 mM sodium pyrophosphate, 1 mM NaF, 1 mM β-glycerophosphate, 0.1 mM sodium orthovanadate, 1 mM PMSF, 2 μg ml−1 leupeptin and 10 μg ml−1 aprotinin. Co-immunoprecipitations were performed by mixing nuclear and cytoplasmic fractions with antibodies and Dynabeads protein G (Life Technology, 10004D) in buffer (50 mM Tris (PH 7.5), 100 mM NaCl, 1.5 mM EGTA and 0.1% Triton X-100) at 4^0C overnight. Dynabeads were washed with Co-IP buffer for 8 times, and immunoprecipitated proteins were analyzed by Western blotting. Antibodies for co-immunoprecipitations were against FOS (Cell Signaling, #2250), NF-κB1 (SCBT, SC-372X), and STAT3 (Cell Signaling, #9139), and antibodies for western blotting were against STAT3 (Cell Signaling, #12640), RELA (Cell Signaling, #8242), and JUNB (Cell Signaling, #3753 and SCBT, SC-8051).

Micronucleus assays. Cells were seeded into 8-well chamber slides (LAB-TEK, #154941), transformed for indicated times, fixed with 4% Formaldehyde, and then stained with DAPI (2 μg/ml) for 5 minutes. The pictures were captured using Michael Widefield inverted Nikon Ti2 fluorescence microscope at Nikon Imaging Center, Harvard Medical School. Micronuclei were counted from 5 random fields of each time point.

ChIP-seq and DNase-seq to define chromatin states. ChIP-seq (7) and DNsae-seq (8) were performed as described previously. Fastq reads were aligned to human reference genome (hg19) using Bowtie (9) allowing up to 2 mismatches. Only the uniquely mappable reads were used for subsequent analyses. For ChIP-seq data for STAT3, NFKB1, JUN, JUNB, FOS, H3K27ac, H3K4me3 and H3K4me1, we used MACS (10) to call peaks with the cutoff P-value < 10^{-8} in at least one sample, using the parameters "macs2 callpeak --llocal 1000000 -g 2.7e9". For ChIP-seq data for H3K27me3, H3K9me3 and

H3K36me3, we used SICER (11) to call peaks with the cutoff E-value > 40, window size 200 bp and gap size 600 bp, which is better for identifying broad read peaks. For DNase-seq data, we used MACS (10) to call peaks with the cutoff *P*-value < 10^{-11} in at least one sample and using the following parameters "macs2 callpeak --llocal 1000000 -g 2.7e9". Chromatin states were defined as promoters (H3K4me3 peaks), active enhancers (H3K27ac peaks, but no H3K4me3), poised enhancers (H3K4me1 peaks, but no H3K27ac or H3K4me3 peaks) or heterochromatin (H3K9me3 or H3K27me3 peaks).

Analyses of transcription factor binding. Transcription factors tend to localize into *cis*-regulatory regions (CRRs) that regulate gene expression. We merged overlapping peaks of all factors to define CRR regions. For each CRR, we measured factor binding levels as Reads per Million (RPM) using ChIP-seq data, and chromatin accessibility based on DNase-seq data. The peak summit of each factor binding site was defined based on MACS (10). In Figure 1B, we plotted the distance between peak summits of paired transcription factors, located in the same CRR. For each factor binding site, we took 50 nt around the peak summits to perform the motif analyses, using the HOMER (12). We used the position weight matrices (PWM) of STAT, NF-κb and AP-1 in HOMER (12). In Figure 1D, we plotted the distribution of STAT, NF-κb and AP1 motifs around peak summits of transcription factors. As the control, we shuffled the nucleotide positions of PWM, and kept the A/T/G/C occurrence frequency of the motifs as the same. We created the shuffled motifs for 50 times for each motif and plotted their occurrence around the factor peak summits. To examine the contribution of factor binding to gene expression, we assigned factor binding peaks to the closest expressed genes within a distance of 200 megabases, summed up the ChIP signal, and calculated the binding level fold-change during transformation.

Transcriptional profiling. RNA was extracted using mRNeasy Mini Kit following the manufacturer's instruction. RNA-seq libraries were prepared using TruSeq Ribo Profile Mammalian Kit (Illumina, RPHMR12126) as per manufacturer's instruction. RNA-seq libraries were sequenced by Harvard Bauer Core Facility using Hiseq 2000. Raw reads were aligned to GENCODE (13) defined transcripts and then human reference genome (hg19) using Tophat (9) allowing up to 2 mismatches. The gene expression levels were indicated as transcript per million (TPM) values. For Figure 4D and 4E, we downloaded

microarray data to from GSE17941 to examine gene expression dynamics during MCF10-ER-Src and fibroblast cell transformation (3). The gene expression values were calculated by the RMA approach using Affymetrix Expression Console Software.

Cancer patient data analyses. The gene expression, genetic and lineage annotations of 1,036 cancer cell lines were downloaded from the Cancer Cell Line Encyclopedia (CCLE) (14). The level 3 data showing clinical annotations and RNA expression of cancer patients were downloaded from TCGA database. The tumor purity estimations were obtained from (15), which estimated tumor purity using 5 different measurements: STIMATE, based on gene expression profiles of 141 immune genes and 141 stromal genes; ABSOLUTE, based on somatic copy-number data; LUMP, based on 44 non-methylated immune-specific CpG sites; IHC, based on image analysis of haematoxylin and eosin stain slides produced by the Nationwide Children's Hospital Biospecimen Core Resource; and the averaged values based on 4 methods above. The levels of infiltrating immune cells obtained from (16). Single-Cell RNA sequencing data in melanoma patients were obtained from (17).

Calculation of cancer inflammation index. The 27 signature genes in the inflammatory loop include genes in IL1/NF-κb pathway (IL1A, IL1B, IL1R1, IL1R2, IL1RAP, IL1RL1, MYD88, IRAK2, NFKB1 and NFKB2), IL6/STAT3 pathways (IL6, LIF, OSMR, JAK2 and STAT3), TNF /AP-1 pathway (TNFSF10, TNFRSF10D, TNFRSF11B, TNFRSF21, ATF3, FOS, FOSL1, FOSL2, JUN and JUNB) and MAP kinases (MAP3K8 and MAP4K4). We calculated Spearman's rank correlation coefficient values between gene pairs using gene expression data from CCLE and TCGA. For Figure 4C, we randomly selected expressed genes and calculated Spearman's correlation as the background distribution. We grouped breast cancer patients based on their genetic subtypes as following: Triple Negative (ER-, PR- and HER2-), Luminal A (ER+, PR+ and HER2-), Luminal B (ER+, PR+ and HER2+), and HER2+ (ER-, PR- and HER2+).

We calculated the inflammatory index based on the log2 expression levels of these genes indicated as E_{ij}, for the gene i in the sample j. Suppose the total gene number to calculate the index is n and the total sample number is m. We first normalized the gene expression levels across all samples to their median values calculated as $N_{ij} = E_{ij} - median(E_{i1}, E_{i2}, \ldots E_{im})$. Then for each sample j, we calculated

the median expression level of signature genes as $S_j=median(N_{1j}, N_{2j}, \ldots N_{nj})$. To obtain the baseline of inflammatory levels across samples, we used the lower 5 percentile of expression levels for each gene across samples as the basal expression as B_i, and calculated the median expression level as M_i. The final inflammatory index for a sample j is calculated as $I_j=S_j+median((M_1-B_1), (M_2-B_2), \ldots, (M_n-B_n))$. The same calculation step is used for calculating the overall expression levels of angiogenesis genes, apoptosis genes, metastasis/migration genes, and STAT3/NF-kb/AP1 target genes in Figures 7 and S8.

Calculation of cancer progression index. We picked 98 genes showing significant positive correlation with the cancer inflammation index across different cancer types (Figure 6B), are linked to "cell migration", "angiogenesis" and "apoptosis", and are not immune/inflammation related based on gene ontology definition. These genes include: ACTC1, ACTN4, ADAM10, ADAMTSL4, AHR, ANXA2, ARHGAP22, C8ORF4, CARD6, CAV1, CFLAR, CIB1, COL15A1, COL1A1, COL4A2, CSRNP1, CYP1B1, CYR61, DAB2, DAP, DFNA5, DLC1, DRAM1, ECE1, EDN1, ELK3, EPAS1, EPHB2, F3, FN1, GADD45A, GADD45B, HBEGF, HIPK3, HMOX1, HSPG2, HTATIP2, IER3, IGFBP3, ILK, ITGA5, ITGAV, JAG1, JUN, LAMB1, LAMC2, LEPR, LGALS1, MYADM, MYH9, MYO1C, NEK6, NFKBIA, NOTCH2, NRP1, NUMB, PEA15, PHLDA1, PHLDA2, PLAU, PLK3, PNPLA6, PPP1R13L, PPP1R15A, PRKCA, PTPRB, PTPRH, RALB, RHBDD1, RRAS2, RTN4, SDCBP, SEMA4B, SERPINB3, SGK1, SH3GLB1, SH3KBP1, SH3RF1, SHB, SHC1, SLC12A6, SQSTM1, SRPX2, STEAP3, STK17A, STK17B, STX4, TGFA, TGFBI, TGFBR2, TMEM214, TSPO, UBE2Z, VEGFA, WARS, WASF2, XAF1 and ZFP36L1. We used the same calculation steps described for the cancer inflammation index to determine the cancer progression index.

Correlation between inflammation levels and drug response. The drug response data of 267 compounds in 1,001 cancer cell lines were obtained from the Genomics of Drug Sensitivity in Cancer (18).

Gene Ontology analyses. The Database for Annotation, Visualization and Integrated Discovery (DAVID) (19) was used for gene ontology analyses.

Data availability. All sequencing data that support the findings of this study have been deposited in the National Cancer for Biotechnology Information Gene Expression Omnibus (GEO) and are accessible through the GEO series accession numbers GSE115597, GSE115598 and GSE115599. All computational codes are available from the authors upon request.

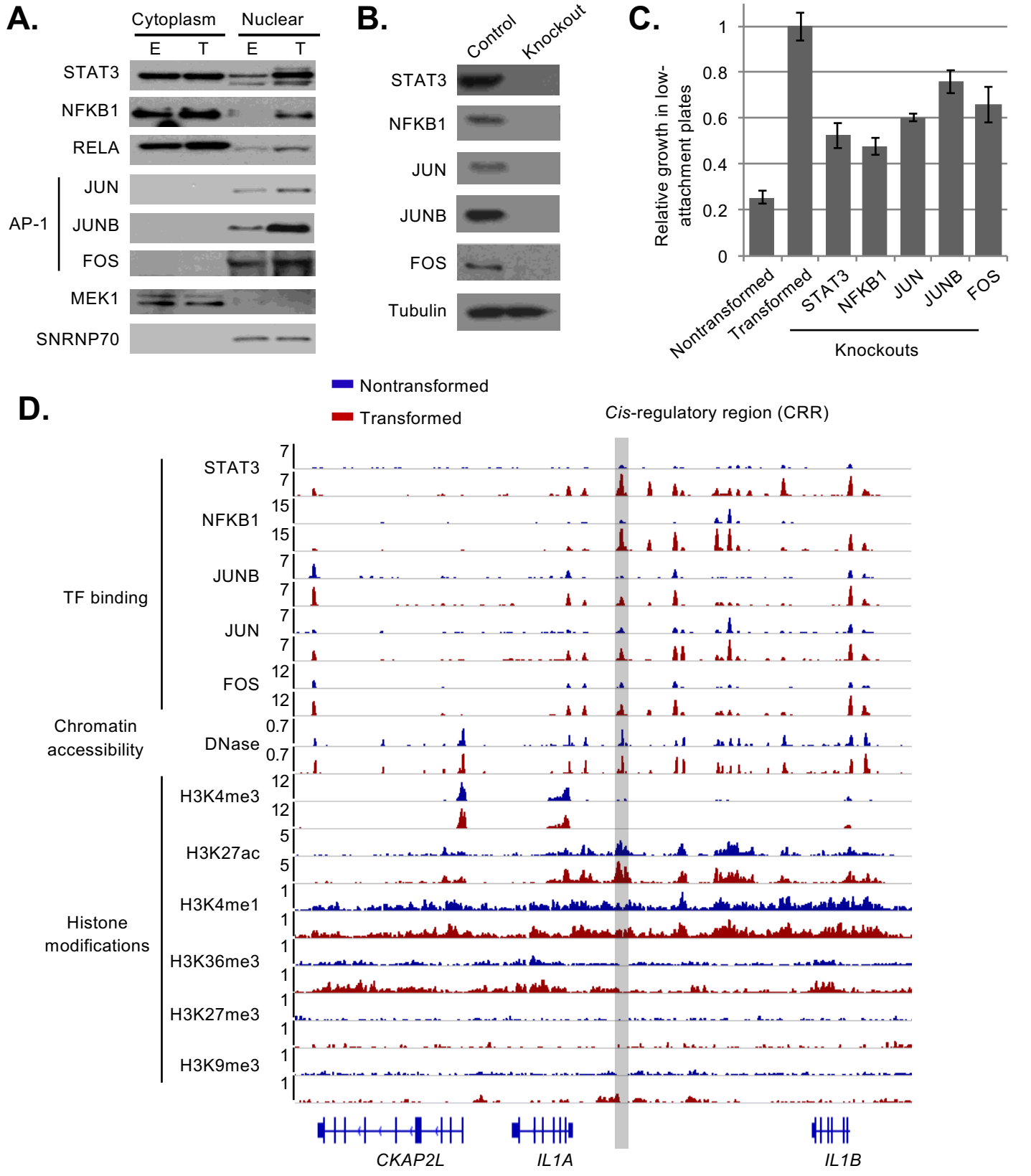


Figure S1. ChIP-seq for STAT3, NFKB1, JUN, JUNB and FOS during MCF10A-ER-Src cell transformation. (*A*) Western blots showing expression of the indicated proteins in the nucleus and cytoplasm during transformation. (*B*) Western blots confirming CRISPR knockouts of the indicated transcription factors. Tubulin was used for the control. (*C*) Cell transformation assays for cells upon factor knockouts and the control. (*D*) A genomic region showing ChIP-seq data for the indicated transcription factors and histone modifications and DNase-seq data for measuring chromatin accessibility before and after transformation.

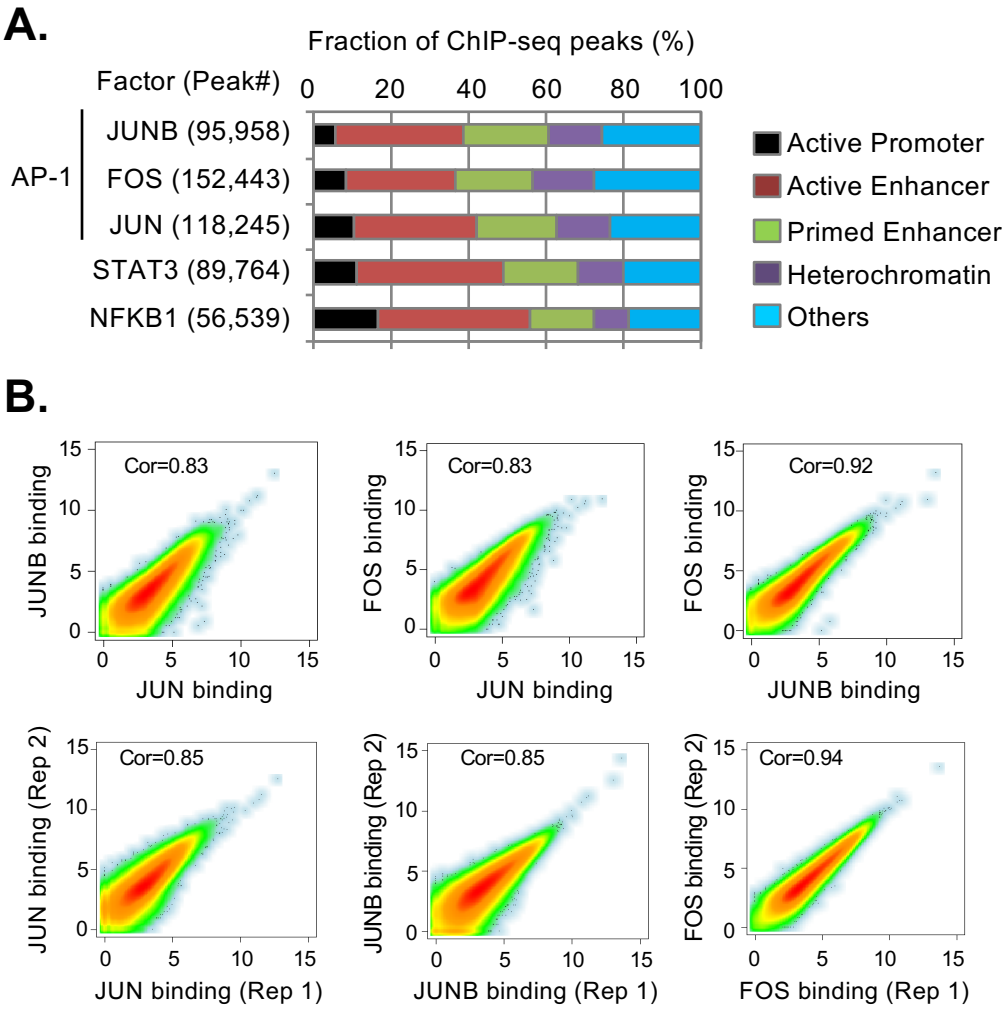
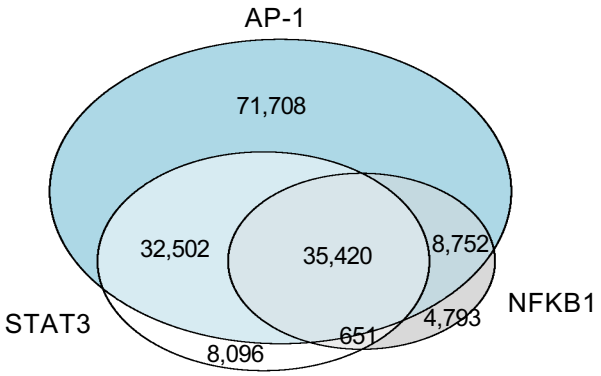
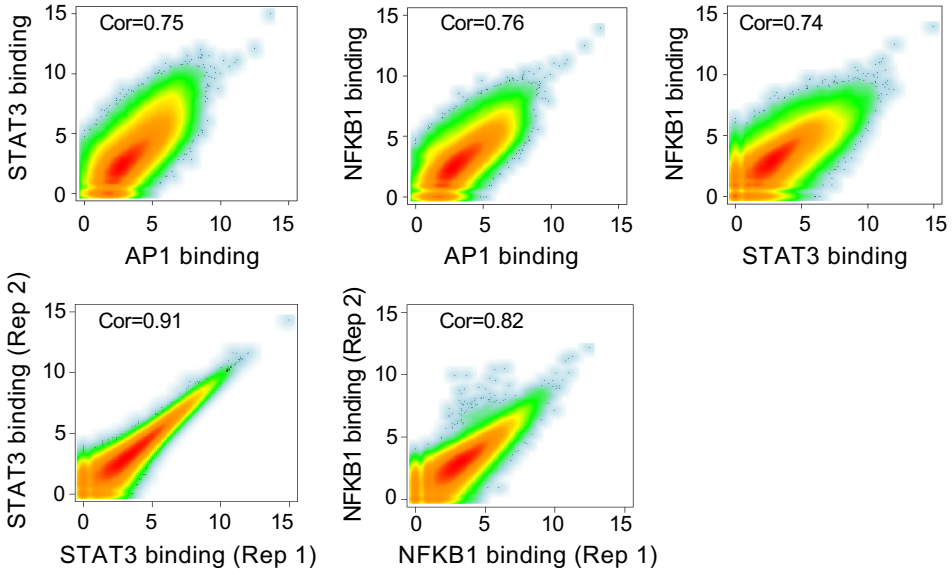


Figure S2. Genomic distribution of STAT3, NF-kB and AP-1 binding sites. (*A*) Genomic locations of binding sites for the indicated transcription factors. Regulatory regions were classified by chromatin status as follows: active promoters (H3K4me3); active enhancers (H3K27ac, but no H3K4me3); primed enhancer (H3K4me1 only); Heterochromatin (H3K9me3 or H3K27me3); others (without histone modification peaks). (*B*) Correlation of JUN, JUNB and FOS binding levels, and correlation of binding levels between biological replicates. We merged overlapping peaks for these AP-1 factors for the analyses. The Pearson correlation coefficiency values are shown.

A.



B.



C.

	AP-1 	STAT 	NF-κB 
STAT3 binding sites	37.8%	24.1%	4.7%
NFKB1 binding sites	30.4%	7.9%	16.0%
AP-1 binding sites	60.3%	6.5%	3.0%
Random sequences	3.2%	4.7%	1.9%

Figure S3. Location analyses and motif analyses of factor binding sites. (*A*) Number of cis-regulatory regions (CRRs) bound by STAT3, NFKB1 and AP-1. (*B*) Correlation of AP-1, STAT3, NFKB1 binding levels, and correlation of binding levels between biological replicates. (*C*) Fractions of factor binding sites with AP-1, STAT, NF-κβ motifs. We picked 10,000 random genomic sequences as the control.

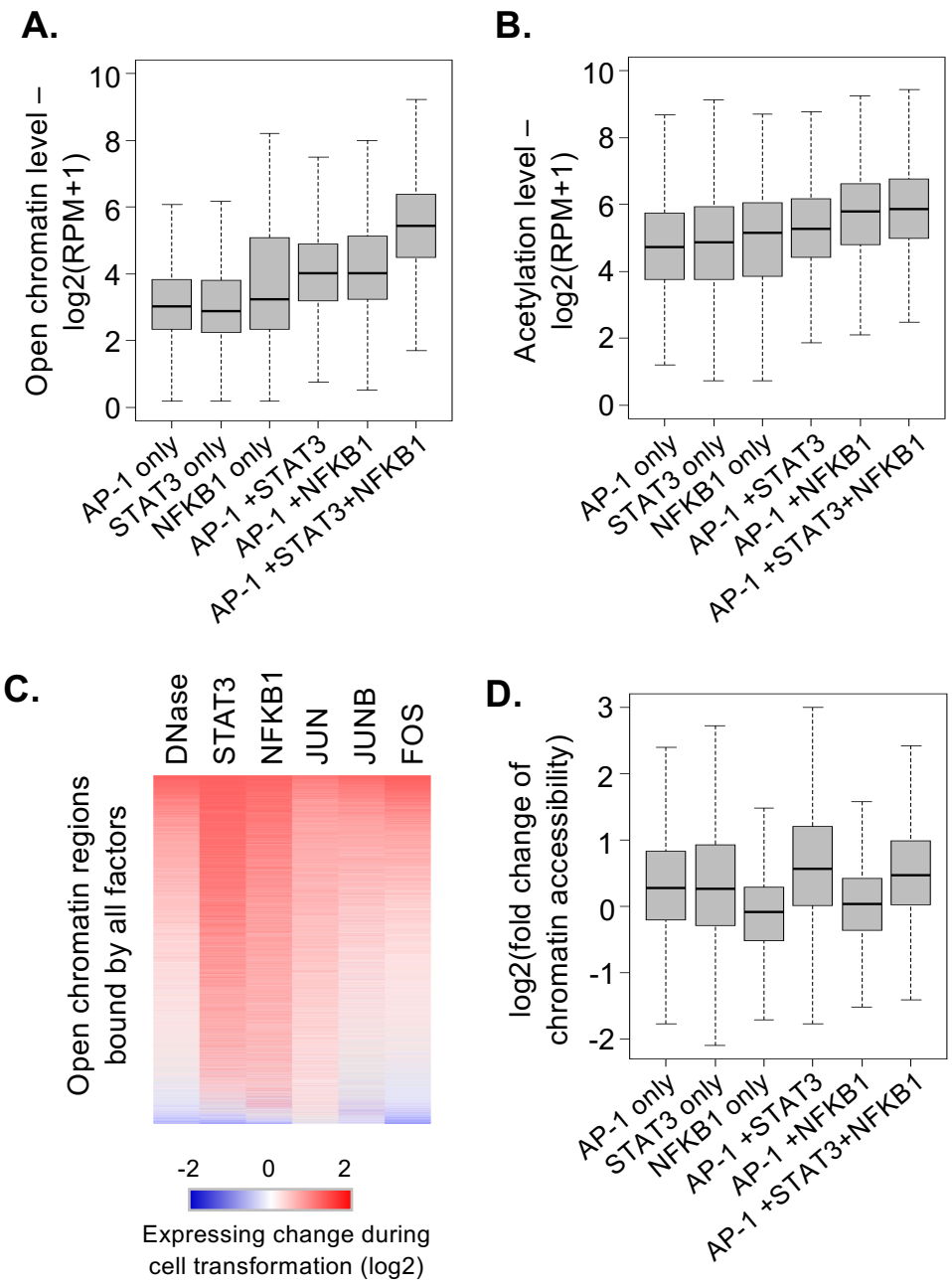


Figure S4. Relationship between factor binding and chromatin accessibility**.** (*A*) Chromatin accessibility (DNase-seq) of regions bound by the indicated factors. (*B*) Histone acetylation (H3K27ac) levels of regions bound by the indicated factors. (*C*) Heatmap showing the dynamic regulation of chromatin accessibility and factor binding levels for open chromatin regions with binding of the indicated factors. (*D*) Differential chromatin accessibilities bound by the indicated factors during transformation.

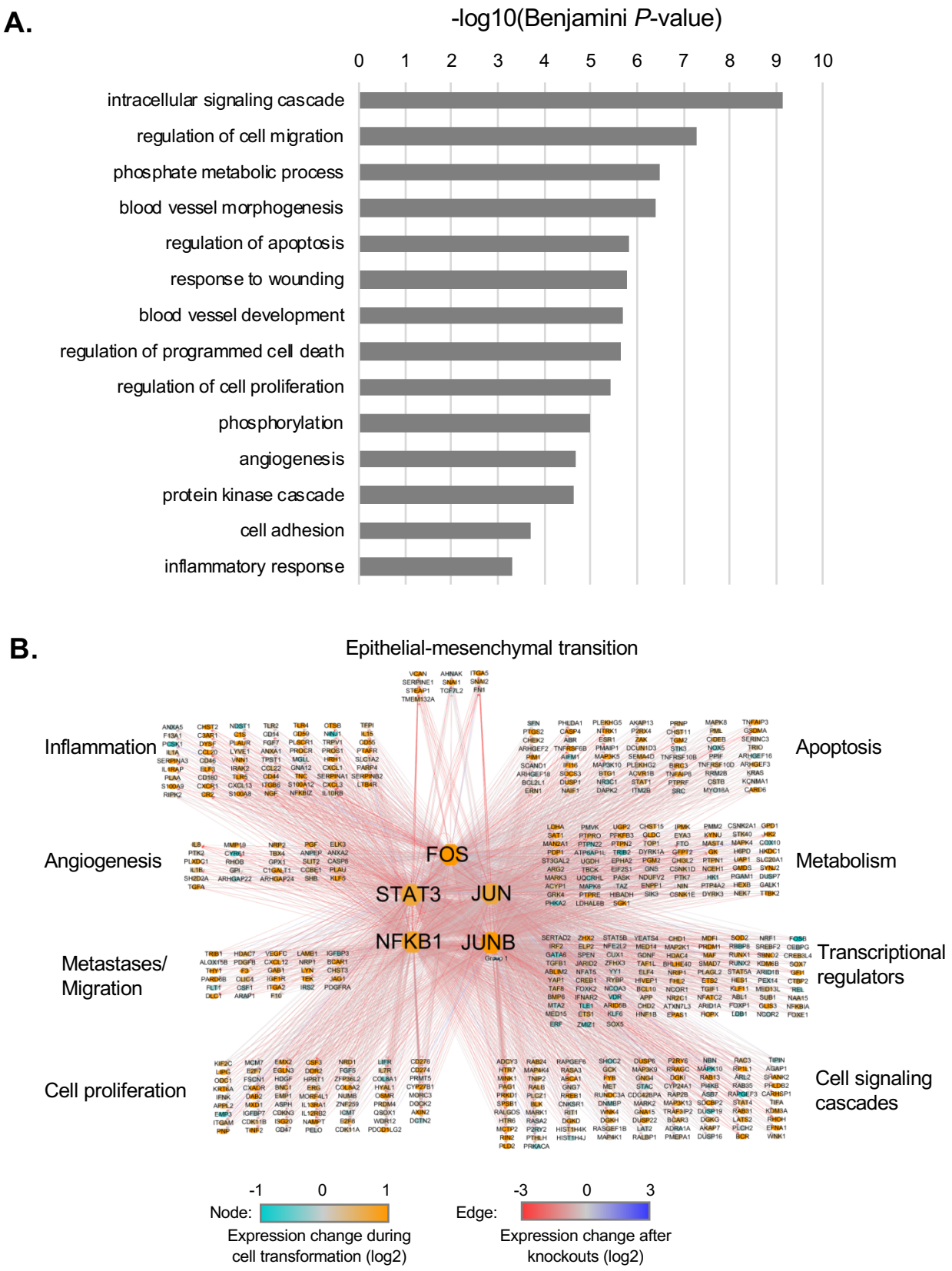


Figure S5. Transcriptional regulatory circuit. (A) Gene ontology analyses. Classes of genes that are common targets of STAT3, NFKB1, JUN, JUNB and FOS, show >1.5 fold increased binding of at least 4 factors during transformation, and are down-regulated upon at least 4 factor knockouts. (*B*) Transcriptional network mediated by STAT3, NFKB1 and AP-1 factors in different oncogenic pathways. The nodes represent factors and target genes in the network. The edges represent direct binding of factors in promoter/enhancer regions of target genes, and gene expression change upon factor knockouts.

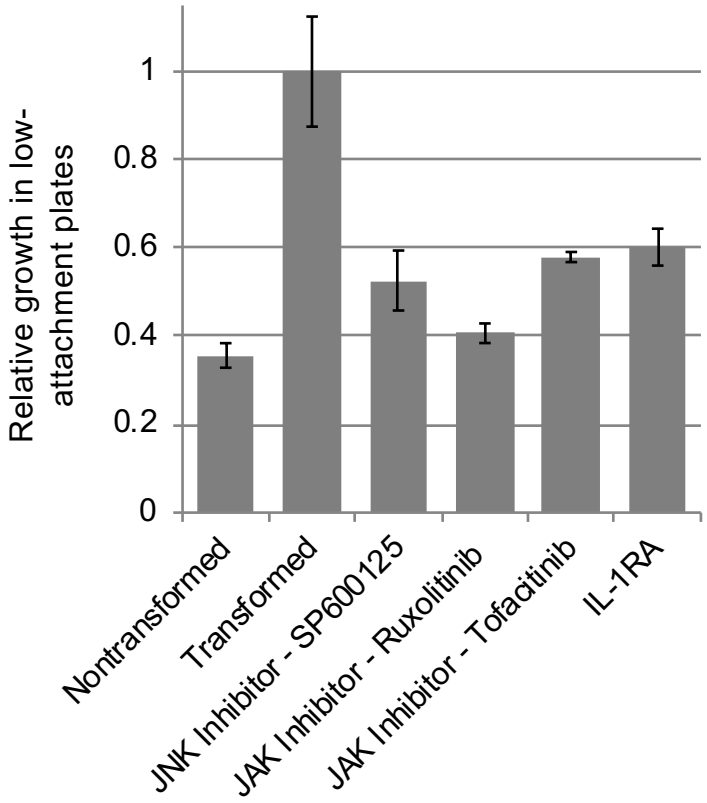


Figure S6. Pharmacological inhibition of JNK, JAK, and IL1 receptor pathways reduces transformation. Transformation efficiencies of cells treated with upon treatment of JNK inhibitor (- 0.3 μM SP600125, 0.3 μM Ruxolitinib, 0.1 μM Tofacitinib, and 30 ng/ml IL1 receptor antagonist.

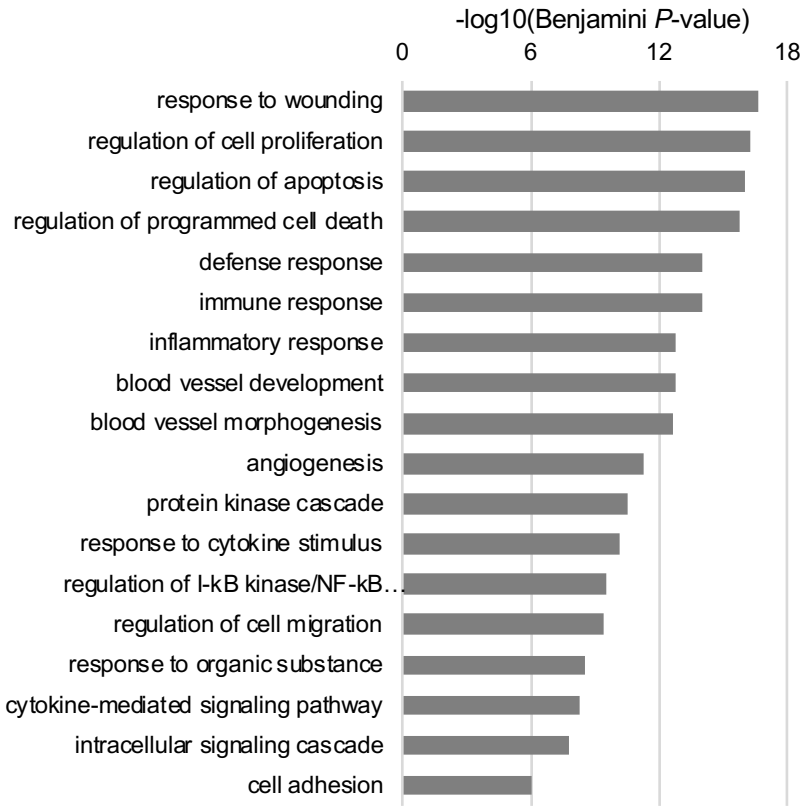


Figure S7. Gene ontology of genes showing significant positive correlation of expression and inflammatory index (median Spearman correlation coefficiencies across cancer types > 0.18).

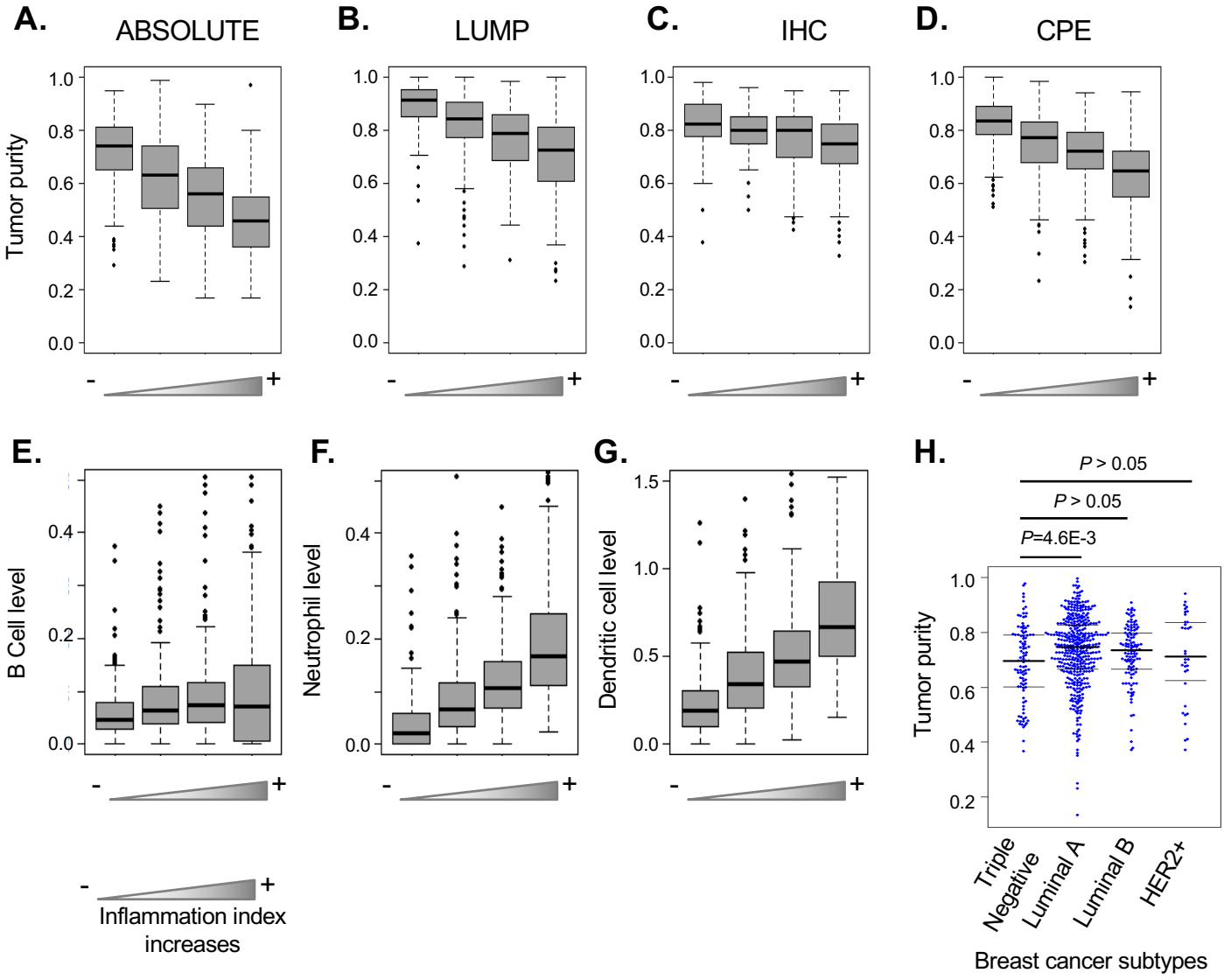


Figure S8. Distribution of the inflammation levels, purity and infatuating immune cells of breast tumors. (A-D) Correlation between tumor purities and the inflammation levels. The tumor purities were estimated using (A) ABSOLUTE, (B) LUMP, (C) IHC, and (D) CPE methods. (E-G) Correlation between inflammation levels and estimated B-cell, neutrophil and dendritic cell levels in the tumors. (H) The tumor purities of breast cancer subtypes.

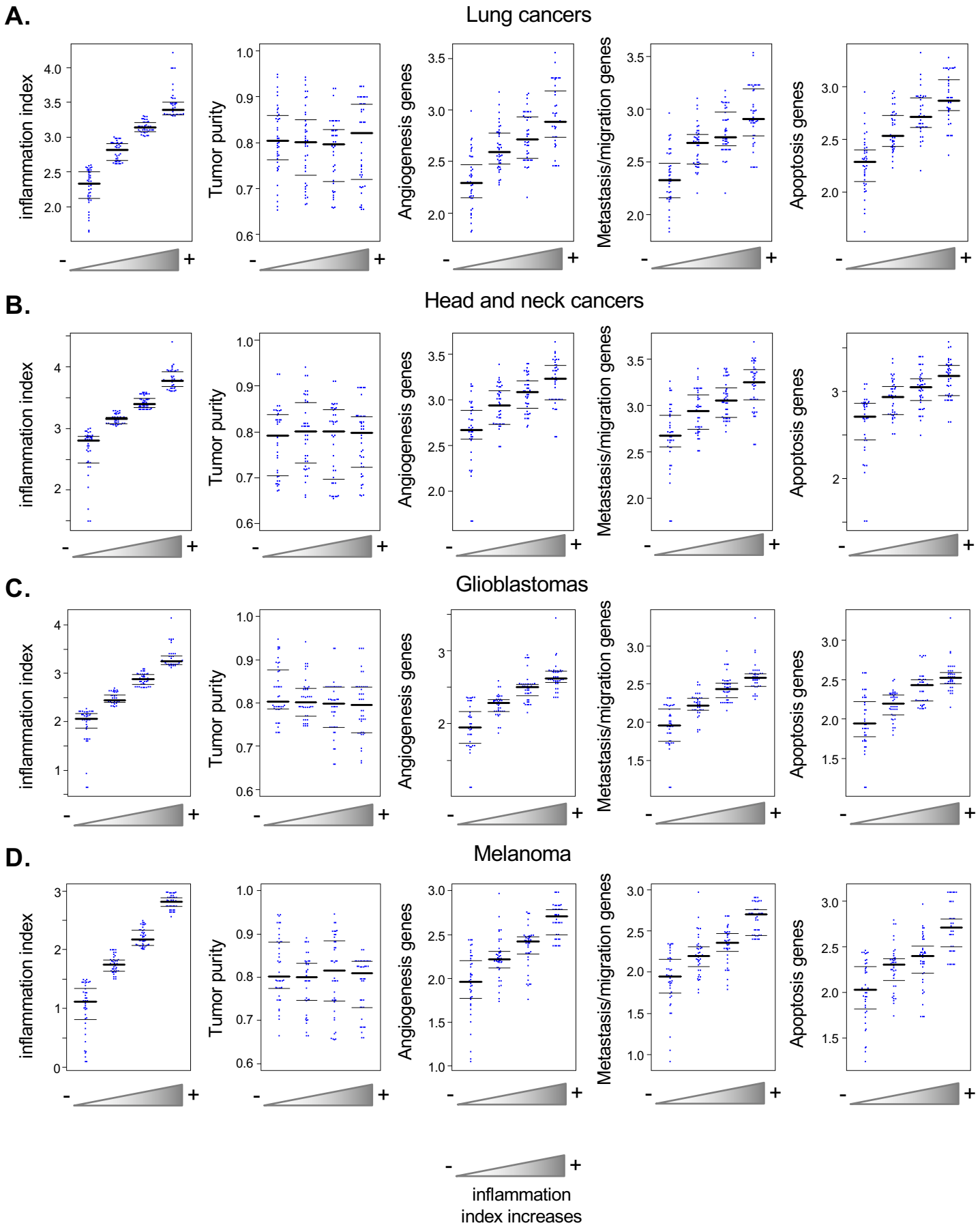
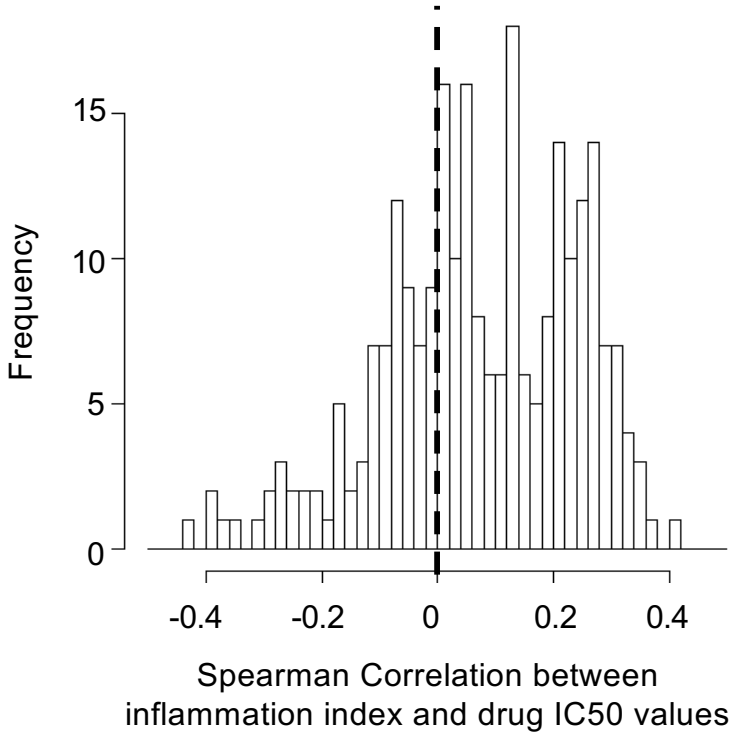


Figure S9. Relationship between tumor purity and inflammation. Randomly picked tumor samples with different inflammation levels and similar purity were examined for expression levels of genes in the indicated oncogenic pathways. Genes related to immune/inflammatory responses were excluded from the analyses.

A.



B.

Drug	Correlation	*P*-value	Drug Target
TL-2-105	0.400	1.85E-26	not defined
Quizartinib	0.375	5.31E-18	FLT3
NPK76-II-72-1	0.358	9.50E-23	PLK3
SB52334	0.358	2.23E-20	ALK5
Amuvatinib	0.344	6.33E-17	KIT, PDGFRA, FLT3
Vorinostat	0.339	1.24E-15	HDAC inhibitor Class I, IIa, IIb, IV
I-BET-762	0.337	1.09E-20	BRD2, BRD3, BRD4
GSK1070916	0.337	1.69E-18	AURKA, AURKC
Tubastatin A	0.332	4.19E-20	HDAC1, HDAC6, HDAC8
YM201636	0.319	1.13E-15	PYKFYVE
Ruxolitinib	0.243	2.06E-07	JAK1, JAK2

Figure S10. The correlation between inflammation levels of cancer cells and drug response. (A) The Spearman Correlation Coefficiency between inflammation levels and IC50 values of drug compounds across 1,001 human cancer cell lines. (B) The drug compounds showing better efficacy in highly inflammatory cancer cells.

REFERENCES

1. Soule HD, *et al*. (1990) Isolation and characterization of a spontaneously immortallized human breast epithelial cell line, MCF10. *Cancer Res* 50:6075-6086.

2. Aziz N, Cherwinski H, & McMahon M (1999) Complementation of defective colony-stimulating factor 1 receptor signaling and mitogenesis by Raf and v-Src. *Mol. Cell. Biol.* 19:1101-1115.

3. Hirsch HA, *et al*. (2010) A transcriptional signature and common gene networks link cancer with lipid metabolism and diverse human diseases. *Cancer Cell* 17:348-361.

4. Iliopoulos D, Hirsch HA, & Struhl K (2009) An epigenetic switch involving NF-kB, lin 28, let-7 microRNA, and IL6 links inflammation to cell transformation. *Cell* 139:693-706.

5. Rotem A, *et al*. (2015) Alternative to the soft-agar assay that permits high-throughput drug and genetic screens for cellular transformation. *Proc Natl Acad Sci USA*.112:5708-5713.

6. He L, *et al*. (2011) alpha-Mannosidase 2C1 attenuates PTEN function in prostate cancer cells. *Nat Commun* 2:307.

7. Fleming JD, *et al*. (2015) STAT3 acts through pre-existing nucleosome-depleted regions bound by FOS during an epigenetic switch linking inflammation to cancer. *Epigenetics Chromatin* 8:7.

8. Thurman RE, *et al*. (2012) The accessible chromatin landscape of the human genome. *Nature* 489:75-82.

9. Langmead B, Trapnell C, Pop M, & Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25.

10. Zhang Y, *et al*. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome biology* 9(9):R137.

11. Zang C, *et al*. (2009) A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* 25:1952-1958.

12. Heinz S, *et al*. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38:576-589.

13. Harrow J, *et al*. (2012) GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 22:1760-1774.

14. Barretina J, *et al*. (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483:603-607.

15. Aran D, Sirota M, & Butte AJ (2015) Systematic pan-cancer analysis of tumour purity. *Nat. Commun.* 6:8971.

16. Li B, *et al*. (2016) Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol* 17(1):174.

17. Tirosh I, *et al*. (2016) Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352:189-196.

18. Iorio F, *et al*. (2016) A landscape of pharmacogenomic interactions in cancer. *Cell* 166(3):740-754.

19. Huang da W, Sherman BT, & Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protoc* 4:44-57.