Uncovering vulnerabilities in triple-negative breast cancer

He, J.

Chapter 5

Integrative analysis of genomic amplification-dependent expression and loss-of-function screen identifies ASAP1 as a driver gene in triple-negative breast cancer progression

Jichao He, Ronan P. McLaughlin, Lambert van der Beek, Sander Canisius, Lodewyk Wessels, John W. M. Martens, John A. Foekens, Yinghui Zhang, Bob van de Water

Manuscript in preparation

Abstract

The genetically heterogeneous triple-negative breast cancer (TNBC) continues to be an intractable disease, due to lack of effective targeted therapies. Gene amplification is a major event in tumorigenesis. Genes with amplification-dependent expression are being explored as therapeutic targets for cancer treatment. In this study, we have applied ADMIRE (Analytical Multi-scale Identification of Recurring Events) analysis and transcript quantification in the TNBC genome across 222 TNBC tumors (n = 118 TCGA and n = 104 Metabric) and identified 148 candidate genes with positive correlation in copy number gain (CNG) and gene expression. siRNA-based loss-of-function screen of the candidate genes has validated EGFR, MYC, ASAP1, IRF2BP2 and CCT5 genes as drivers promoting proliferation in different TNBC cells. MYC, ASAP1, IRF2BP2 and CCT5 display frequent CNG and concurrent expression over 2173 breast cancer tumors (cBioPortal dataset). More frequently are MYC and ASAP1 amplified in TNBC tumors (> 30%, n = 320). In particular, high expression of ASAP1, the ADP-ribosylation factor (Arf) GTPase-activating protein, is significantly related to poor metastatic relapse free survival of TNBC patients (n = 257, bc-GenExMiner). Furthermore, we have revealed that silencing of ASAP1 modulates numerous cytokine and apoptosis signaling components, such as IL1B, TRAF1, AIFM2 and MAP3K11 that are clinically relevant to survival outcomes of TNBC patients. ASAP1 has been reported to promote invasion and metastasis in various cancer cells. Our findings that ASAP1 is an amplification-dependent TNBC driver gene promoting TNBC cell proliferation, functioning upstream apoptosis components and correlating to clinical outcomes of TNBC patients, support ASAP1 as a potential actionable target for TNBC treatment.

Keywords

Recurrent copy number gain; Driver gene; Whole transcriptome sequencing; Triple-negative breast cancer; ASAP1

Background

Triple-negative breast cancer (TNBC) represents a particularly proliferative and aggressive subtype of breast cancer, associated with large size of tumors, high mitotic rate of tumor cells, high tumor grade and metastasis, causing poor prognosis and high mortality rate of patients [107]. TNBC constitutes 15-20% of breast cancer, being clinically characterized by the lack of expression of estrogen receptor (ER), progesterone receptor (PR) and the absence of amplification of human epidermal growth factor receptor 2 (HER2), which are the known drivers of other breast cancer types [232]. The absence of actionable targets defined in TNBC cells leads to clinical failure of targeted therapies in TNBC patients. Cytotoxic chemotherapy remains the conventional systemic treatment for TNBC patients, resulting in adverse effects and unfavorable outcomes [77, 233]. Identification of actionable targets for TNBC treatment is a continuous effort.

Gene copy number alterations (CNAs) are a hallmark in the cancer genome [49]. Gains or losses of gene copy are important somatic genomic aberrations contributing to tumorigenesis [106]. Changes in gene copy number result in corresponding changes in expression of the affected genes, causing phenotypes [234]. Copy number gains (CNGs), increasing from the two DNA copies present in normal diploid genome, sometimes to several hundred copies (known as amplification), are observed to frequently occur on cancer driver genes [235, 236]. Oncogenic driver genes with increase in DNA copy number and expression have been identified and explored as potential drug targets for targeted therapies in cancer [237]. For instance, the *HER2* gene, which is amplified in ~30% of primary breast cancers [238], has been proven as an actionable target for trastuzumab antibody targeted therapy and lapatinib inhibitor targeted therapy treating patients with HER2-amplified breast cancer [239]. Therefore, identification of CNG-driven genes and their amplification-dependent overexpression provides opportunities for discovering potential cancer driver genes as therapeutic targets for therapy-refractory cancer.

A number of studies have demonstrated the genomic heterogeneity in TNBCs, being dominated by substantial mutational burdens, including CNAs and genomic rearrangements [106, 240]. CNA genomic profiling based on separate TNBC sample groups have reported the TNBC-related recurrent CNAs on various chromosome regions [241]. Yet, integrative analysis of CNG frequency and CNG-driven gene expression in the TNBC genome is limited.

In this study, we applied the ADMIRE (Analytical Multi-scale Identification of Recurring Events) algorithm to identify candidate driver genes that are frequently amplified by recurrent CNGs, in correlation with their RNA expression levels, in 222 triple-negative tumors from TCGA (n=118) [19] and Metabric (n=104) [232] datasets. As a result, 148 genes were identified, with a significant and positive correlation in their gene amplification and expression. These amplification-driven genes were subsequently validated by loss-of-function screen for their biological function in TNBC cell proliferation,

followed by assessment of their expression and amplification in broad breast cancer cell lines and tumors, and evaluation of their clinical relevance using multiple large public breast cancer datasets [242, 243]. Consequently, we characterized the known oncogenes MYC and EGFR and the novel candidate genes ASAP1, IRF2BP2 and CCT5 as cancer drivers in promoting proliferation of TNBC cells. MYC and ASAP1 were observed to be more frequently amplified and highly expressed in TNBC than non-TNBC tumors. Specifically, high expression of ASAP1, an ADP-ribosylation factor (Arf) GTPase-activation protein regulating cell motility and invasiveness [244], is significantly relevant to poor metastatic relapse-free survival (MRFS) of patients with TNBC tumors, not non-TNBC tumors. Transcriptome analysis further revealed that ASAP1 regulates various cytokine and apoptosis signaling components that are significantly associated with TNBC prognosis. Our work discovered ASAP1 as an amplification-dependent gene driving TNBC proliferation, survival and progression, supporting the potentiality of ASAP1 as a therapeutic target for the treatment of TNBC.

Methods

Cell culture

Human TNBC cell lines BT549, Hs578T and SUM149PT were cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum, 25 U/mL penicillin and 25 µg/mL streptomycin in a humidified incubator at 37°C with 5% CO_2.

Selection of amplification-dependent candidate driver genes in TNBC genome by integrated AMDIRE copy number region analysis and transcript expression quantification

ADMIRE analysis [245] was performed to detect genomic regions with recurrent copy number gains across 222 triple-negative tumors (n=118 from TCGA; n=104 from Metabric). Segmented copy number profiles for both TCGA [19] and the discovery set of Metabric [232] were obtained and used as input for the ADMIRE analysis. ADMIRE was configured to control its false discovery rate at 0.01. The recurrently altered copy number regions identified by ADMIRE were filtered in order to enrich regions most likely to harbor driver genes. Importantly, ADMIRE regions may be nested within larger regions, where the higher nesting levels correspond to more focal and more frequently altered regions. Our filtering only kept the regions detected at the highest nesting level. In addition, regions were kept only if they spanned at least one, but no more than 100 genes. Next, the genes contained in those regions were identified and selected as candidate driver genes, if their mRNA expression profile showed a positive correlation with their copy number. For this, the RNA-Seq data for TCGA analyzed using RSEM transcript quantification and the microarray expression data for Metabric were applied. The correlation was tested using Spearman correlation, where the Log-ratio copy number estimates were correlated with

the expression values. Correction for multiple testing was performed by controlling the false discovery rate at 0.1. The ADMIRE analysis and the subsequent filtering steps were performed separately for TCGA and Metabric, after which the resulting gene lists were merged.

siRNA-mediated loss-of-function screen

The primary screen was carried out by use of siGENOME Human SMARTpool siRNAs (GE Dharmacon, Lafayette, CO, USA) targeting 148 ADMIRE candidate driver genes. In the validation screen, SMARTpool siRNA and single siRNA_1, _2, _3 and _4 that comprise the SMARTpool mix were used to validate each candidate hit. Cells were seeded overnight in 96-well plate at an optimized density for BT549 (8,000 cells/96-well), Hs578T (8,000 cells/96-well) and SUM149PT (10,000 cells/96-well), and transfected with 50 nM siRNA by transfection reagent INTERFERin (Polyplus-Transfection SA, Illkirch-Graffenstaden, France) according to the manufacturer's instructions. We used a pool of 720 kinase siRNAs at stock concentration of 1 µM in our laboratory as a negative control (siCtrl), this has no significant effect on expression of any single kinase genes; siRNA against KIF11 was used as positive control. The medium was refreshed 24 h post-transfection and TNBC cells were transfected for 2 days and proliferated for 4 days under indicated condition. SRB colorimetric assay was used as read-out for cell proliferation.

SRB proliferation assay

A sulforhodamine B (SRB) colorimetric assay was used to measure total amount of proteins indicative of cell proliferation, as previously described [125].

Real time RT-qPCR assay

RNA was isolated from TNBC cells, which were transfected with corresponding siRNA for 72 h, using RNeasy (Qiagen). cDNA was generated from 400 ng total RNA, using RNeasy Plus Kit from Qiagen. Real-time RT-qPCR was performed in triplicate, using the SYBRGreen PCR MasterMix (Applied Biosystems) on a 7900HT fast real-time PCR system (Applied Biosystems). The primer sequences used were: forward *CAGCCGGCGCTTCCC*, reverse *ATCAGAAAACGACCGGGACC* (human ASAP1); forward *CTGGTAAAGTGGATATTGTTGCCAT*, reverse *TGGAATCATATTGGAACATGTAAACC* (human GAPDH). Relative mRNA levels after correction for GAPDH control mRNA were expressed using the $2^{-\Delta\Delta CT}$ method.

DNA copy number alteration and mRNA expression profiling of candidate genes in breast cancer cell lines and tumors

DNA copy number data of candidate genes in 20 TNBC cell lines were obtained from online resources [246]. Log2-based RNA expression profiles of candidate genes in 52 breast cancer cell lines was retrieved from our own established RNA-Seq data. Copy number

alterations (CNAs) and mRNA expression of candidate hits in 2173 breast tumors were obtained from dataset "METABRIC, Nature 2012 & Nat Commun 2016" in cBioPortal (*http://www.cbioportal.org/*), an open-access resource for interactively exploring multidimensional cancer genomics [243] and the largest dataset with available DNA copy number (n = 2173) and mRNA expression (n = 1904) profiles. The 2173 breast tumor samples were filtrated by the immunohistochemistry (IHC) status of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2), resulting in 320 triple-negative tumors.

Human whole transcriptome analysis of ASAP1 silencing effect

TNBC cells were seeded into 96-well plate and transfected with optimal SMARTpool siRNA targeting ASAP1 (siASAP1), and siRNA control (siCtrl), as described above. The experiment was performed in biological triplicate. 72 h later, cells were washed with PBS and lysed in 80 µl 1× BioSpyder lysis buffer. Lysates were frozen at − 80 °C and shipped to BioSpyder technologies on dry ice for human whole transcriptome targeted RNA sequencing TempO-Seq analysis. Expression data for 21,111 transcripts were generated (BioSpyder Technologies, Inc., Carlsbad, CA, United States). Normalization and differential expression analysis were performed using DESeq2 package. Specifically, each siASAP1 condition was paired with the corresponding control siCtrl and the counts for each sample were normalized using the DESeq2 estimateSizeFactors function. Differential expression of each treatment relative to its respective control was measured using the Wald test. With baseMean of counts < 10 filtered, genes that were regulated by siASAP1 with significance (p value < 0.05 and absolute Log2 FC > 1) were considered significantly differentially expressed genes (DEGs).

Gene functional enrichment analysis of ASAP1-regulated genes

The DEGs, up- or down-regulated by silencing ASAP1, were uploaded to Metascape [167], an oriented resource combining functional enrichment, interactome analysis, gene annotation and membership research to leverage over 40 independent knowledgebases within one integrated portal (http://metascape.org). Pathway and process enrichment analysis was carried out with ontology sources of KEGG pathway, GO Biological Processes, Reactome Gene Sets, Canonical Pathways and CORUM. All genes in the genome were used as the enrichment background. Terms with p-value < 0.01, a minimum count of 3, and an enrichment factor (a ratio between the observed counts and the counts expected by chance) > 1.5 were collected and grouped into clusters based on their membership similarities. To further capture the relationship between the terms of pathways and processes, network of enriched terms were visualized using Cytoscape, where each node represents an enriched term and is colored by its cluster ID. Furthermore, if network contains between 3 and 500 proteins, protein-protein interaction enrichment analysis was

carried out with BioGrid, InWeb_IM and OmniPath to identify densely connected network components, presented in MCODE (Molecular Complex Detection) network node.

Assessment of clinical relevance of candidate genes to survival prognosis of breast cancer patients

The functionally validated candidate driver genes were further evaluated using Kaplan-Meier analysis for their relation to overall survival (OS) of 1981 breast cancer patients according to their gene expression [243]. Microarray DNA expression results from BrCa Gene-Expression Miner (bc-GenExMiner) were used to classify prognostic association of ASAP1 expression levels with metastatic relapse free survival (MRFS) of 257 TNBC patients (n=257) and ER+ BC patients (n=2519) using "optimal" splitting criterion [242]. The ASAP1-regulated DEGs involved in cytokine, lipid metabolism and apoptosis pathways were assessed for their relation to relapse-free survival of TNBC patients by Kaplan-Meier plotter using "Auto select best cutoff" [247]. Mean expression of DEGs was used to assess their prognostic significance.

Statistical analysis

Pearson correlation analysis was performed using GraphPad Prism 7. Statistical analysis of all experimental data was performed using two-way ANOVA (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). Data were expressed as mean ± SEM. Significance was set at $p < 0.05$. The hierarchical clustering in heatmap was performed using CRAN pheatmap package in RStudio (version 0.99.887).

Results

ADMIRE analysis of TNBC genomes identifies TNBC candidate driver genes with recurrent copy number gain and correlated expression

Recurrent CNAs have been recognized as the result of natural selection in tumor evolution, and hence the recurrently altered regions are likely to harbor cancer driver genes [248]. In order to identify candidate driver genes for TNBC, we applied ADMIRE, a robust algorithm for the discovery of broad and focal recurring events [245], to detect genomic regions with frequent CNAs in a set of TNBC tumors (n = 118 for TCGA; n = 104 for Metabric). Aggregated DNA copy number profiles (Figure 1A) assisted in pinpointing recurrently altered regions (Figure 1B). Genes contained in the focal regions were further assessed with copy number and expression correlation analysis, as exemplified for the proto-oncogene MYC and the RNA genes CASC8 (Cancer Susceptibility 8) and PVT1 (MYC activator) (Figure 1C). Genes showing positive correlation in mRNA expression levels and copy numbers were filtered out as candidate drivers (Figure 1D). The ADMIRE analysis and the subsequent filtering steps were performed separately for TCGA and Metabric cohorts. Subsequently, 148 genes were selected as candidate driver genes for TNBC (Additional file

1: Table S1). Next, functional enrichment analysis displayed the significant implication of

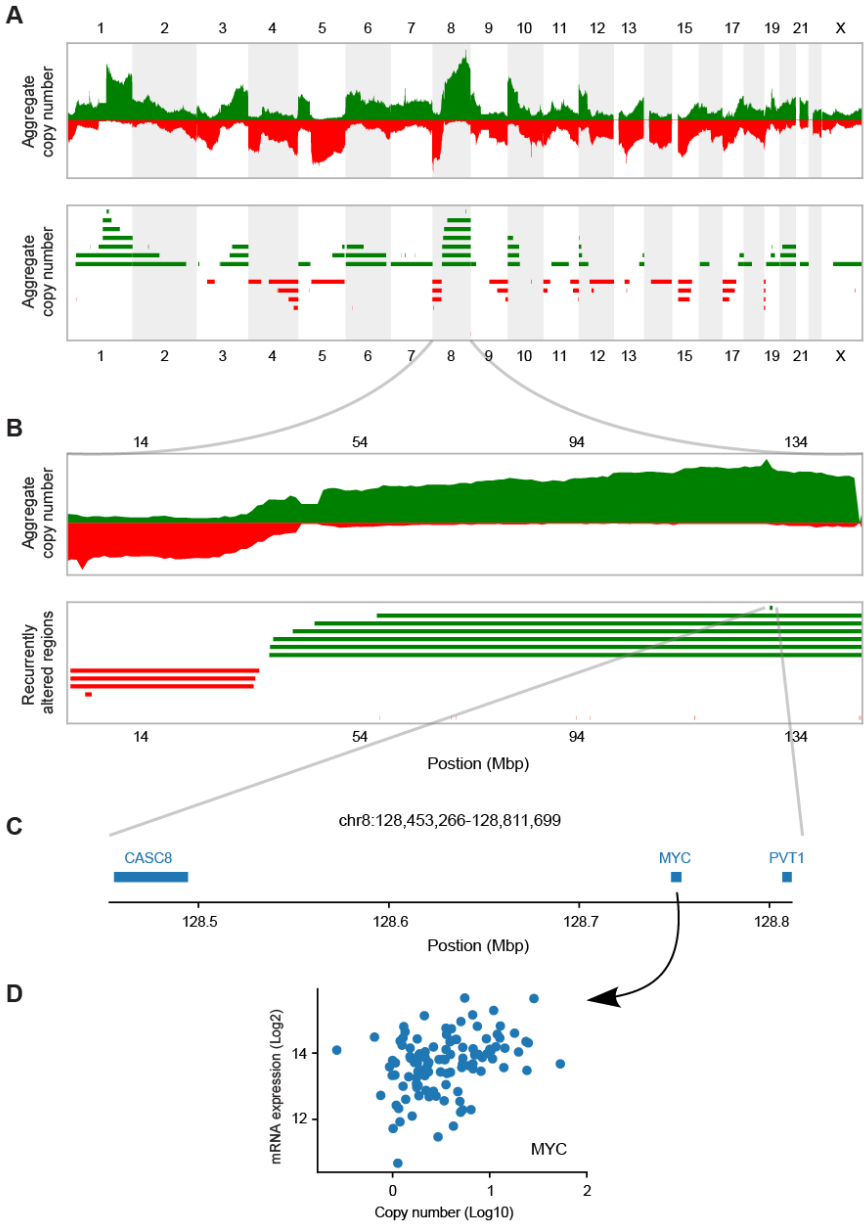


Figure 1. Schematic selection of candidate driver genes with increases in copy number and expression in TNBC genome. (A) Discovery of genomic regions with recurrent copy number alterations using ADMIRE analysis. The top panel shows the aggregate copy number profile across 222 triple-negative breast tumors (n = 118 for TCGA; n = 104 for Metabric). The bottom panel shows the significant recurrent copy number regions, with gains in green and losses in red. **(B)** Zoomed in fragment of panel A, focusing on chromosome 8. The bottom panel reveals the small focal recurrent copy number gain on 8q for further analysis. **(C)** The genes contained in the focal region identified in panel B. **(D)** Scatterplot exemplifying the positive correlation of MYC gene expression with its copy number in TNBC patients.

the 148 genes in 18 KEGG pathways (Additional file 2: Figure S1). The 148 gene set was enriched in cancer-related pathways, including Central carbon metabolism in cancer, Proteoglycans in cancer, Pathways in cancer, Melanoma, Prostate cancer and Endometrial cancer. Moreover, these candidate drivers were implicated in various oncogenic signaling pathways that control cell growth, survival and motility, such as ErbB, PI3K-Akt, Ras, MAPK and Wnt pathways. Altogether, our ADMIRE genomics approach collected 148 candidate TNBC drivers that are frequently amplified by recurrent CNGs in the TNBC genome and significantly enriched in pathways that fuel cancer progression.

Loss-of-function screen validates the TNBC candidate driver genes in TNBC cell proliferation

As a first step, we assessed the biological function of the 148 candidate genes in control of TNBC cell proliferation by siRNA-mediated loss of function screen in two TNBC cell lines, mesenchymal-like BT549 and basal-like SUM149T. High reproducibility was achieved for the duplicate screens in BT549 (r = 0.9212) and SUM149PT (r = 0.9127) (Figure 2A). Genes whose silencing controlled proliferation less than 60% were considered as candidate hits with significance (Figure 2B). Consequently, 41 primary hits were screened for BT549 and 20 for SUM149PT cell line (Figure 2B-C). In total, 46 primary hits were selected, of which 15 were common (Figure 2D; Additional file 3: Table S2).

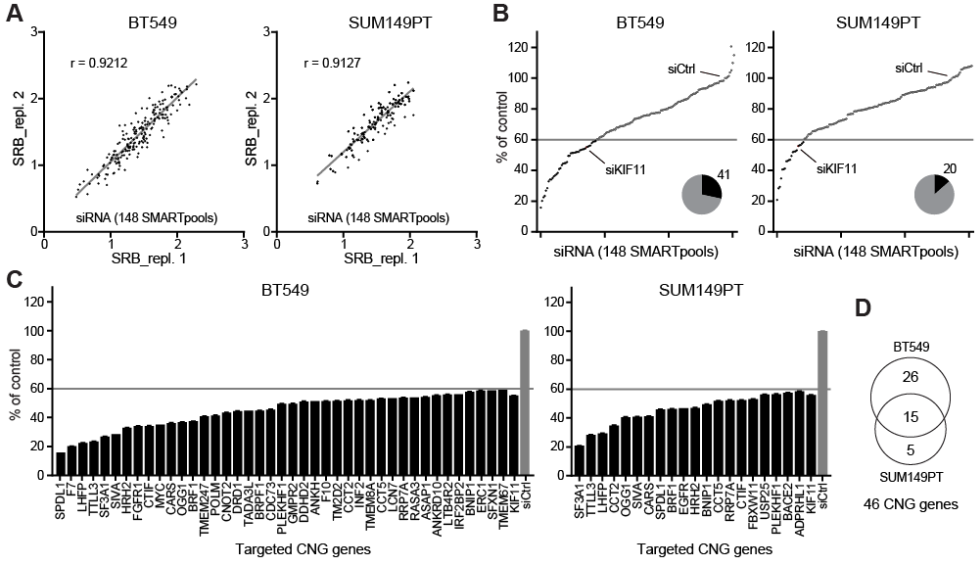
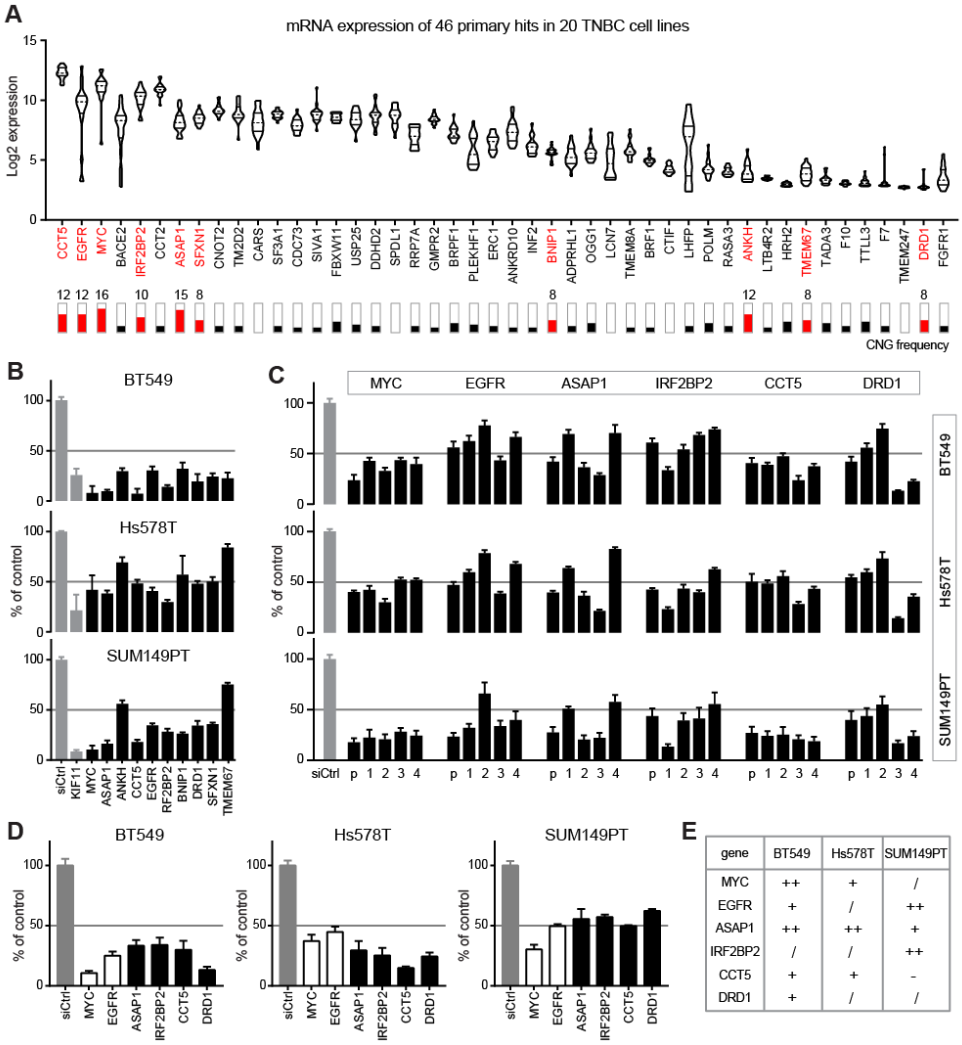


Figure 2. siRNA-mediated loss-of-function screen of candidate driver genes in TNBC cells. (A) Replicate siRNA screens of candidate driver genes in TNBC cell lines BT549 and SUM149T. siRNA silencing effect of candidate gene on cell proliferation was assessed 96 h after transfection and presented with sulforhodamine B (SRB) colorimetric raw values. **(B)** Normalized percentage of proliferation control by siRNA silencing. The number of genes (black) whose silencing led to > 40% proliferative inhibition was indicated in the pie chart, as primary hits.

(C) Ranking and listing of the primary hits significantly controlling proliferation of BT549 and SUM149PT TNBC cell lines. siRNA targeting KIF11, positive functional control; siCtrl, non-targeting siRNA control. siRNA silencing effect on proliferation was relative to siCtrl. Error bars indicate variation of screen replicates. **(D)** Overlap primary hits in BT549 and SUM149T TNBC cell lines.



gene	BT549	Hs578T	SUM149PT
MYC	++	+	/
EGFR	+	/	++
ASAP1	++	++	+
IRF2BP2	/	/	++
CCT5	+	+	-
DRD1	+	/	/

Figure 3. Validation of candidate driver genes with concurrent copy number gain (CNG) and overexpression in TNBC cells. (A) mRNA expression of 46 primary hits in 20 TNBC cell lines. Violin plot indicates Log2 mRNA expression level of 46 primary hits in 20 TNBC cell lines retrieved from RNA-Seq analysis. Bars indicate CNG frequency of the hits in 20 TNBC cell lines. Genes with frequent CNG in ≥ 8/20 TNBC cell lines were marked in red. **(B)** siRNA validation of primary candidate hits with high frequent CNG in BT549, Hs578T and SUM149PT TNBC cell lines. SMARTpool siRNAs were used to target each hit. KIF11 was taken as positive control. **(C)** siRNA deconvolution validation of six candidate driver hits. The effects of SMARTpool (p) siRNA and single siRNA_1, _2, _3 and _4 on hits were compared for their proliferation control (%) in the TNBC cell lines. **(D)** Percentage of control proliferation (%) by optimized SMARTpool siRNAs targeting the six candidate hits. **(E)** CNA of the six

candidate driver genes in TNBC cell panel. "++", high CNG; "+", CNG; "/", no copy number alteration; "-", copy number loss.

Next, we examined the CNG frequency of the 46 primary hits in 20 TNBC cell lines, representative for diverse molecular subtypes of TNBC. Ten candidate hits were found with CNG frequency across ≥ 8 TNBC cell lines (Figure 3A, lower panel). The CNG recurrence for MYC was found in 16/20 of TNBC cell lines, ASAP1 in 15/20, ANKH, CCT5 and EGFR in 12/20, IRF2BP2 in 10/20, BNIP1, DRD1, SFXN1 and TMEM67 in 8/20, respectively. We also evaluated the mRNA expression level of the 46 primary hits in the 20 TNBC cell line panel in our established RNA-Seq data (Figure 3A, upper panel). Given the nature of heterogeneity of TNBC, the correlation of expression and copy number of the primary hits varied among cell lines.

To further validate the 10 candidate hits for their function in proliferation, we performed siRNA silencing in three TNBC cell lines, BT549, SUM149T and one more mesenchymal-like cell line Hs578T. Silencing of 6 hits, ASAP1, CCT5, IRF2BP2, DRD1 and two known oncogenes MYC and EGFR, potently inhibited proliferation (> 50%) in all three cell lines (Figure 3B), suggesting their driving role in TNBC cell proliferation. Deconvolution siRNA screen confirmed the effect of single siRNAs (≥ 2/4) on the proliferation-driving hits, mostly achieving > 50% of proliferation inhibition, ruling out the off-targeting effect of pooled siRNAs (Figure 3C). The optimized pooled siRNA silencing further validated the function of the driver hits in proliferation control (> 50%) (Figure 3D). Of note, the inhibitory effects on cell proliferation by silencing these six hits were, in general, concordant with CNA status in these cell lines (Figure 3E).

Collectively, our RNAi-based functional screen validated the ADMIRE-identified candidate driver genes and defined the role of MYC, EGFR, ASAP1, IRF2BP2, CCT5 and DRD1 in controlling proliferation of TNBC cells.

Frequent amplification of ASAP1 in TNBC is significantly relevant to poor clinical outcome

We next sought to address the correlation between CNA frequencies and expression levels of MYC, EGFR, ASAP1, IRF2BP2, CCT5 and DRD1 in broad breast cancer cell lines and breast tumors and the clinical relevance of the genes to patients with breast cancer. RNA-Seq analyses of the transcriptome from our 52 BC cell lines demonstrated that the Log2-based mRNA expression levels of MYC, EGFR, ASAP1 and CCT5 were higher in TNBC than non-TNBC cell lines (Additional file 4: Table S3), while similar high IRF2BP2 and low DRD1 were expressed in both cell types (Figure 4A). The difference of EGFR expression in TNBC and non-TNBC cells was of significance. Next, we obtained Z score-based mRNA expression data for the hits in 1904 breast cancer (BC) tumors from cBioPortal database [232, 243, 249]. MYC, EGFR, ASAP1 and IRF2BP2 were highly expressed in TNBC, compared to non-TNBC

tumors (Figure 4B). CCT5 expression was found lower in TNBC than non-TNBC breast

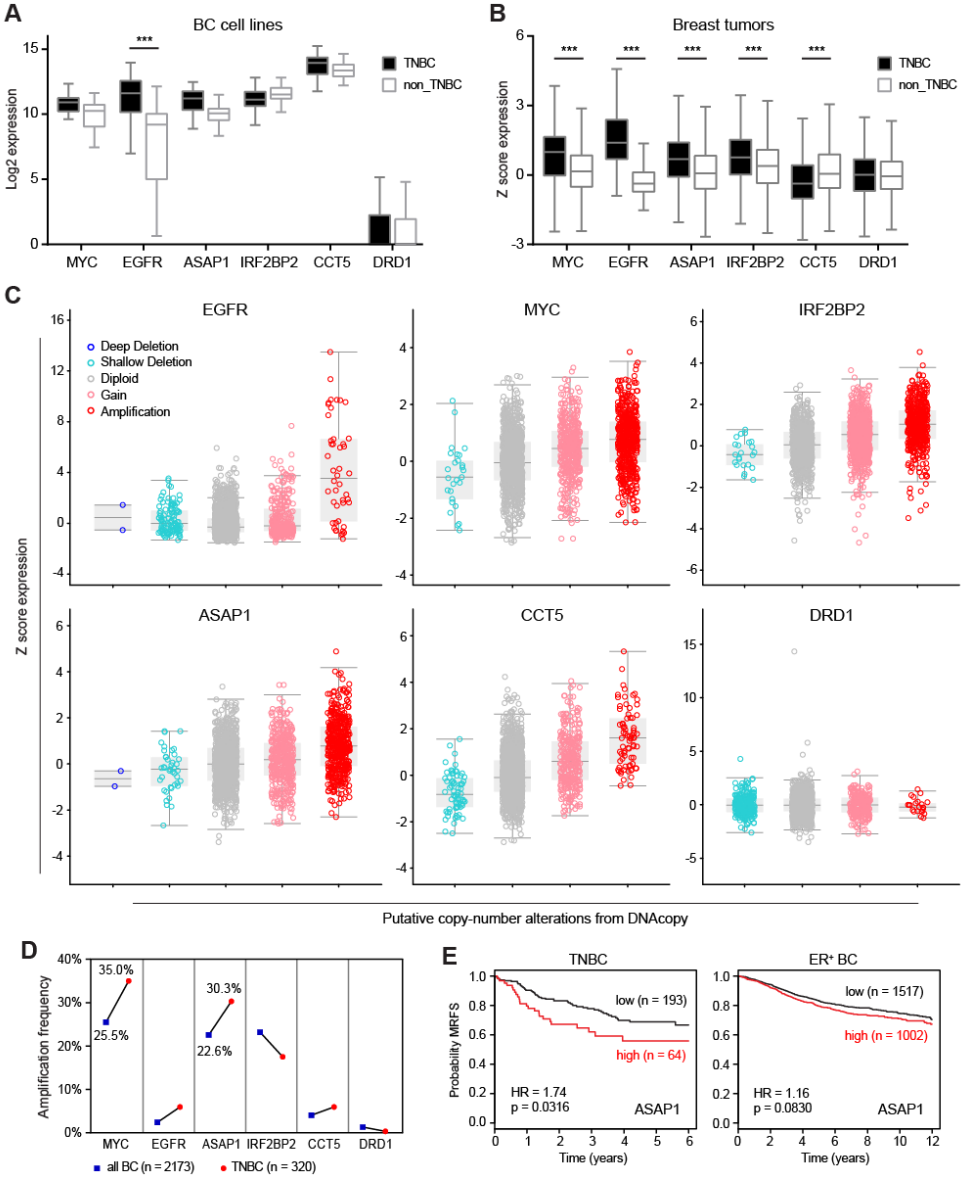


Figure 4. ASAP1 amplification and overexpression in TNBC in association with poor clinical outcome. (A) Log2-based mRNA expression of the six candidate hits in 52 breast cancer (BC) cell lines. Log2 values were obtained from established RNA-Seq data (two-way ANOVA *** $p < 0.001$). **(B)** Z score-based mRNA expression of the hits in 1904 BC tumors. Data were retrieved from the dataset "METABRIC, Nature 2012 & Nat Commun 2016" in cBioPortal dataset. **(C)** Correlation between CNA and gene expression of the candidate hits in the cohort of 1904 BC tumors. Different copy-number amplifications (shallow deletion, diploid, gain, amplification) are presented per gene. **(D)** Amplification frequency of the candidate hits in the cohort of 2173 BC tumors. Data were retrieved from the dataset "METABRIC, Nature 2012 & Nat Commun 2016" in cBioPortal dataset. **(E)** Metastatic relapse-

free survival (MRFS) Kaplan-Meier (KM) curve of ASAP1 in TNBC (n = 257) and ER^+ BC (n = 2519) cohorts analyzed by bc-GenExMiner v4.2.

tumors, and DRD1 showed no significant difference. We further illuminated the correlation between CNAs and expression levels of the genes in the cohort of 1904 BC tumors. While CNA events in deep or shallow depletion rarely or less occurred, EGFR, MYC, IRF2BP2, ASAP1 and CCT5 (except for DRD1) often underwent copy gain and amplification, acquiring CNA-driven expression in the breast tumors (Figure 4C). Moreover, MYC, ASAP1 and IRF2BP2 were amplified in > 20% of all 2173 BC tumors (Figure 4D). Particularly, MYC and ASAP1 amplifications emerged more frequently in TNBC tumors (> 30%), suggesting TNBC subtype-related MYC and ASAP1 amplifications. Prognostic analysis demonstrated the association of the CNA-driven hits with overall survival of the cohort of 1981 BC patients [232, 249], indicating the significant implication of MYC and ASAP1 in poor disease outcomes (Additional file 5: Figure S2). More specifically, high ASAP1 expression was revealed to be related to worse MRFS of patients with TNBC (n=257), but not ER^+ BC (n=2519) (Figure 4E), as assessed in the cohorts of BC patients [242]. Together, we characterized MYC and ASAP1 as CNA-driven genes with frequent amplification and concurrent high expression in BC tumors. The CNA-driven amplification and expression of ASAP1 exhibited significant clinical impact particularly on TNBC progression.

Transcriptomic analysis of the impact of ASAP1 depletion on gene expression in TNBC cells

Amplification or deletion of a gene copy may affect the expression of genes located outside the amplified/deleted region itself via indirect mechanisms [250, 251]. The ASAP1 gene, located at chromosome 8q24.1, encodes an Arf GTPase-activating protein (GAP) that induces hydrolysis of GTP bound to Arf proteins. ASAP1 has been reported to be involved in signal transduction, membrane trafficking and cytoskeleton remodeling [252] and promote proliferative, invasive and metastatic phenotypes of various cancer cells [244, 253]. Whether CNA-driven ASAP1 amplification and overexpression influence gene expression at a genome-wide level in cancer cells is not addressed. To this end, we performed TempO-Seq-based targeted whole genome RNA sequencing in the three TNBC cell lines BT549, Hs578T and SUM149PT that harbor ASAP1 amplification and high expression (Figure 3E). The cells were transfected with siRNA targeting ASAP1 (siASAP1) or non-targeting siCtrl in biological triplicate, respectively. The transcriptome TempO-Seq assays were run at a read depth of 6.8M per sample (Additional file 6: Figure S3A), achieving reproducibility with Pearson r values over 0.95 (Additional file 6: Figure S3B). Principal components analysis (PCA) of global changes in gene expression clustered various gene expression patterns across cell lines and transfections (Additional file 6: Figure S3C). Using DESeq2 package in R, the Log2 normalized transcriptome profiles displayed the differential effects of ASAP1

depletion on gene expression in the TNBC cells (Figure 5A). Silencing of ASAP1 by siASAP1 decreased > 45% of ASAP1 mRNA levels in the TNBC cells (Figure 5B, left panel), warranting the effective knockdown of ASAP1 itself (Log2 FC -0.7 to -1.5) in the TempO-Seq transcriptome panels (Figure 5B, right panel). Genes with 2-fold changes (absolute Log2 FC ≥ 1) in down- or up-regulation (p-value < 0.05) were selected in BT549 (311 down / 495 up), Hs578T (133 down / 117 up) and SUM149PT (500 down / 401 up) cells, respectively (Figure 5C). Venn diagrams extracted differentially expressed genes (DEGs) that were significantly down- or up-regulated by ASAP1 depletion in the TNBC cell lines (Figure 5D; Additional file 7: Table S4). Consequently, 95 DEGs were downregulated, and 79 DEGs upregulated in ≥ 2/3 of the TNBC cell lines, in total 174 DEGs, which were considered as common DEGs that were susceptible to the depletion of the amplification-dependent ASAP1.

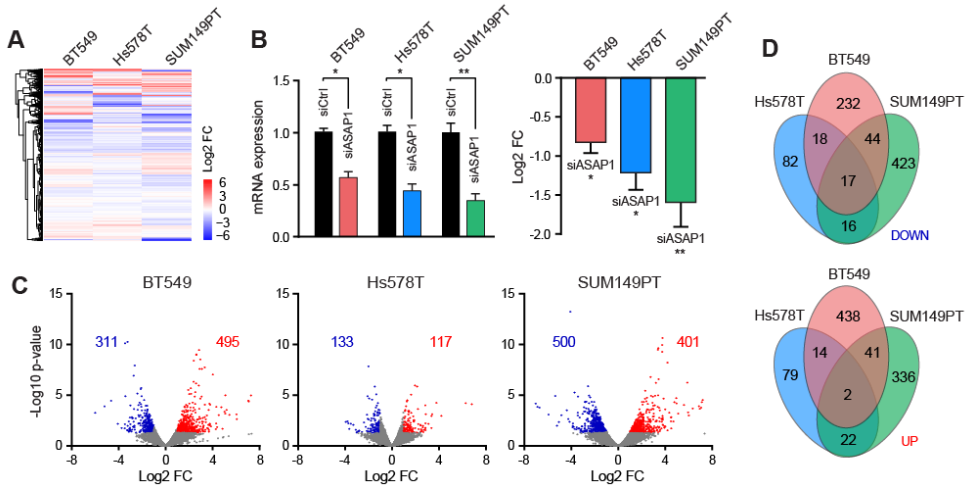


Figure 5. Targeted whole transcriptome analysis of ASAP1 depletion-induced transcription reprogramming in TNBC cells. (A) Transcriptome expression profiling in BT549, Hs578T and SUM149PT TNBC cells after siRNA-mediated depletion of ASAP1 (siASAP1). Log2 fold change (Log2 FC), siASAP1 versus siCtrl. **(B)** Targeting effect of siASAP1 on ASAP1 gene expression. Left panel, knockdown efficiency assessment by RT-qPCR (two-way ANOVA * $p < 0.05$, ** $p < 0.01$). TNBC cells were transfected with optimized SMARTpool siRNAs for 72 h. GAPDH was used as internal reference. Right panel, Log2 FC of ASAP1 upon knockdown. **(C)** Volcano plot of differentially expressed genes (DEGs) in ASAP1-depleted BT549, Hs578T and SUM149PT cells. The red and blue dots indicate down- and up-regulated DEGs, respectively, with p-value < 0.05 and absolute Log2 FC > 1. **(D)** Venn diagram of down- and up-regulated DEGs in BT549, Hs578T and SUM149PT TNBC cell lines. Upper panel, down-regulated DEGs; lower panel, upregulated DEGs. Common DEGs denote DEGs popping-up in at least two cell lines.

ASAP1 regulates cytokine and apoptosis signaling components that are associated with TNBC prognosis

Next, to evaluate the biological functions of 174 common DEGs that were susceptible to ASAP1 depletion, we conducted Metascape Pathway and Process Enrichment Analysis integrating the gene ontology sources, including GO Biological Process, KEGG pathway,

Reactome Gene Sets, Canonical Pathways and CORUM [167]. Top 20 clusters were defined with their representative enriched terms (Figure 6A, left; Additional file 8: Table S5), including cytokine signaling pathways (cytokine signaling in immune system, response to interleukin-1, TNF-signaling pathway and regulation of inflammatory response), metabolic processes (regulation of lipid metabolic process and regulation of sulfur metabolic process), apoptosis signaling pathways (regulation of apoptosis signaling pathway and positive regulation of cell death), MAPK pathways (regulation of p38 MAPK cascade and MAPK signaling pathway), cell cycle arrest and P53 downstream pathway. Furthermore, network enrichment captured the interactions between the 20 clusters, as visualized using Cytoscape (Figure 6A, right). Strikingly, among the 20 clusters, the cytokine signaling in immune system, regulation of lipid metabolic process and regulation of apoptotic signaling pathway were most significantly enriched. Protein-protein interaction clustering algorithm identified neighborhoods within the networks where the ASAP1-regulated genes were densely connected, such as ANXA1, C3, CXCL1, CXCL2 and CXCL8 node (involved in immune response), IMPDH1, PSMC4 and RAN node (involved in nucleotide and protein metabolism), and APEX1, MCM6 and RBL2 node (involved in cell cycle G1/S phase transition) (Figure 6B). These results revealed the novel and essential biological functions of ASAP1 in multiple molecular pathways.

Around 45% of ASAP1-regulated genes (78/174) were involved in cytokine signaling pathways (Additional file 9: Table S6), being interactive in cytokine signaling in immune system, response to interleukin-1, TNF-signaling pathway and regulation of inflammatory response. In regulation of apoptosis signaling and cell death pathways, 33 genes were modulated by ASAP1 depletion. Unsupervised hierarchical heatmap classified the 21 positive and 20 negative regulators in cytokine signaling pathways, 5 positive and 10 negative regulators in lipid metabolism pathway, and 8 positive and 9 negative regulators in apoptosis signaling pathways (Figure 6C), in the BT549, Hs578T and SUM149PT TNBC cell lines. As cytokine signaling pathways, e.g. TNF-signaling pathway, mediate extrinsic/intrinsic apoptosis [254], crosstalk might exist in the ASAP1-regulated cytokine and apoptosis signaling pathways. Indeed, among 33 ASAP1-regulated apoptosis genes, 23 genes (~70%) were interactive in cytokine signaling pathways (Additional file 9: Table S6), further stressing the implication of ASAP1 as a driver gene in cell survival and growth. Crosstalk was also observed between ASAP1-regulated genes in lipid metabolic process (17/31, ~50%) and cytokine signaling (Additional file 9: Table S6). To explore the association between these ASAP1-regulated cytokine, lipid metabolism and apoptosis genes and the survival of patients with TNBC, we performed Kaplan-Meier (KM) survival analysis using KM plotter [247]. The results exhibited that, based on the mean expression of the selected genes in TNBC tumors, a low expression level of the negative regulators, which were downregulated under ASAP1 depletion, predicted longer relapse-free survival of TNBC patients with statistical significance (logrank p = 0.00015, HR = 2.76) (Figure 6D).

In contrast, low expression of positive regulators, which were upregulated by ASAP1 depletion, implied significantly worse prognosis (logrank p = 0.0092, HR = 0.47).

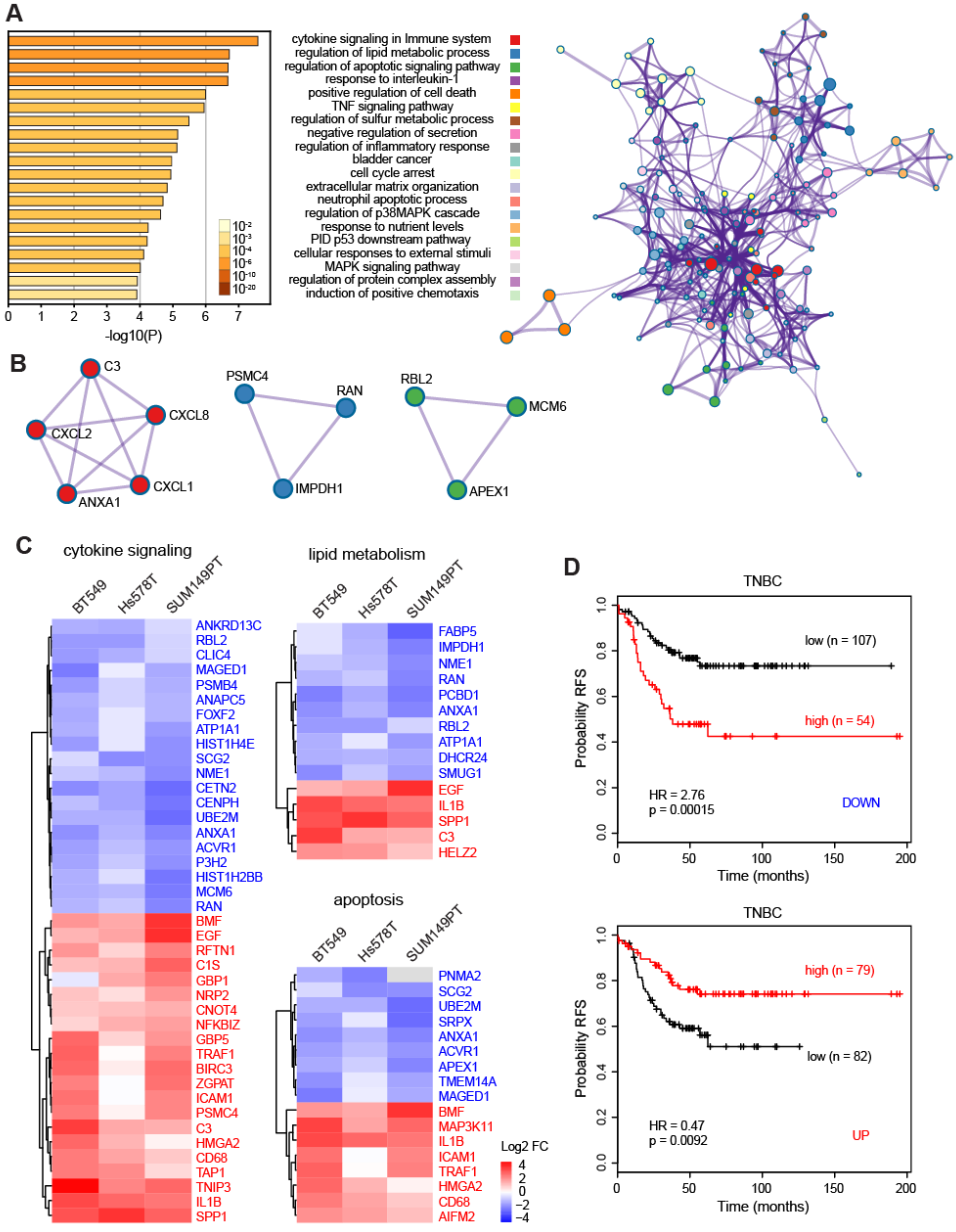


Figure 6. Metascape functional enrichment analysis and TNBC clinical relevance of ASAP1-regulated DEGs. (A) Top 20 clusters with their representative enriched term across input of 174 ASAP1-regulated DEGs. Left panel, heatmap of the 20 enriched terms. One term per cluster, colored by p-values. Log10(P) is the p-value in log base 10. Right panel, network of the 20 enriched terms, colored by cluster ID, where nodes that share the same cluster ID are typically close to each other. **(B)** Representative Molecular Complex Detection (MCODE) network nodes, showing the ASAP1-regulated DEGs densely connected. **(C)** Log2 FC clustering of DEGs involved in

cytokine signaling, lipid metabolism and apoptosis pathways. Blue, negative regulators; red, positive regulators. **(D)** Relapse-free survival (RFS) KM curves of the DEGs from panel C in patients with TNBC. Mean expression of positive DEGs (DOWN) or negative DEGs (UP) was used.

Collectively, our data suggested that amplified ASAP1 is a key driver in TNBC progression by negatively regulating cell death pathways. The ASAP1-regulated genes in cytokine, lipid metabolism and apoptosis signaling pathways were clinically relevant to TNBC relapse-free survival.

Discussion

TNBC is characterized by its inter-tumoral heterogeneity. Based on gene expression signatures, TNBC has been further classified into more than six molecular subtypes [8], Yet, no effective molecular targeted therapies are currently available in the clinic for this type of breast cancer. Although TNBC demonstrates substantial genetic alterations [255], only two genes, *TP53* and *PIK3CA*, have been found with mutation frequency in ~10% of TNBC patient tumors [106], indicating other driver mutations involved in TNBC progression. While amplification frequently occurs in cancer genomes, amplified genes are not always overexpressed [256]. A recent study has applied integrative analysis combining gene expression, miRNA and copy number variation genomic profiles for TNBC patients (n=137) from TCGA to reclassify TNBC inter-tumoral heterogeneity [257]. As overexpression is a requisite for amplified genes to function as drivers in cancer, we attempted to identify TNBC candidate driver genes by integrating DNA copy number change and mRNA expression omics data across 222 TNBC cases (TCGA dataset n=118 and Metabric n=104). Using recurrent event calling algorithm ADMIRE analysis and transcript expression quantification, we consequently identified 148 genes with frequent focal CNGs and concurrent high expression in the cohort of TNBC tumors. Among these genes several are well-known oncogenes with involvement in breast cancer, e.g. MYC, EGFR, CCNE1 and FGFR1, consistent to other studies [19, 106, 258]. We also identified many novel candidates, such as TMEM67, ANKH, BNIP1, CCT5, ASAP1 and IRF2BP2. KEGG pathways enrichment displayed the implication of the genes mainly in cancer-related or oncogenic signaling pathways. Our integrated genomic analysis reveals frequent CNGs and amplification as driver mutations in the TNBC genome. This large group of 148 candidate genes, displaying positive correlation in recurrent DNA CNGs and high RNA expression levels in the TNBC genome, may represent potential drivers during the course of TNBC progression.

Our current siRNA-based proliferation screen validated several candidate genes promoting TNBC cell proliferative phenotype, including the known oncogenes MYC and EGFR and the novel candidates ASAP1, IRF2BP2, CCT5 and DRD1, indicating the implication of amplification-dependent drivers in TNBC proliferation. These genes (except for DRD1) acquired CNG/amplification and high expression in breast tumors. In particular,

amplification of MYC and ASAP1 is highly frequent in TNBC (35% and 30.3%, respectively) compared to non-TNBC tumors (25.5% and 22.6%, respectively). The amplification-based MYC oncogene has been found in various solid tumors, including kidney and colorectal cancers and breast cancer subtypes, including TNBC [47, 259], indicating MYC amplification as a frequent driver mutation in cancer. A recent study has revealed that MYC alterations are mutually exclusive with PIK3CA, PTEN, APC, or BRAF alterations [260], suggesting MYC amplification as a distinct driver mutation in the cancer genome. The novel candidate, ASAP1, has also been reported to be frequently amplified, accompanied by enhanced expression, in different types of cancers, including pancreatic ductal adenocarcinoma, prostate cancer and melanoma [261, 262], exceptionally in primary breast cancer where overexpression of ASAP1 was described to be independent of ASAP1 amplification [263]. These early results, though not reported for breast cancer, underline the impact of ASAP1 amplification over the cancer genome. Here, our results initially interpret that, similar to other cancer types, the aggressive breast cancer subtype TNBC acquires driver mutations, such as amplification-dependent overexpression of MYC and ASAP1, to progress, in spite of genetic heterogeneity. Of significance for ASAP1, we showed that knockdown of ASAP1 inhibits cell proliferation in different TNBC cell lines and high expression of ASAP1 is associated with poor metastatic relapse-free survival of patients with TNBC but not ER+ breast cancer, compelling the role of ASAP1 in TNBC progression, promoting not only proliferation but also metastasis.

In this study, we demonstrated for the first time the effect of ASAP1 on transcriptomic regulation in TNBC cells. ASAP1 (also named DDEF1 or AMAP1) was identified on the basis of Arf activity as a phospholipid-dependent Arf GTPase-activating protein and was found to bind to and be phosphorylated by Src family proteins and focal adhesion kinase (FAK) and to associate with focal adhesions [264, 265]. ASAP1 has been shown to promote cell proliferation and invasion in different cancer cells, including lung, colorectal, prostate and breast cancer cells [244, 261, 266]. High ASAP1 expression level was found in and was required for invadopodia formation in invasive MDA-MB-231 TNBC breast cancer cells, and interfering ASAP1-mediated protein complex inhibited metastasis of MDA-MB-231 derived xenografts [263]. Overexpression of ASAP1 has been reported to be associated with poor metastasis-free survival in prostate, colorectal cancer and ovarian cancer and malignant phenotypes of primary breast cancers [244, 263, 267]. We showed that depletion of ASAP1 in different TNBC cells led to reprogramming of gene expression mainly in cytokine and apoptosis interactive signaling pathways, by upregulating positive regulators and downregulating negative regulators of the pathways. For instance, silencing ASAP1 downregulated the components in the identified mitotic cell cycle G1/S phase transition network node, including the anti-apoptotic APEX1 that is abnormally expressed in numerous human solid tumors and positively correlated with cancer progression [268], the proliferation marker MCM6 that is predictive for poor prognosis in breast cancer [269], and

the direct AKT target RBL2 [270]. In addition, expression of genes involved in lipid metabolic process was also vulnerable to ASAP1 depletion, suggesting the novel role of ASAP1 in metabolism-related tumorigenesis, as lipid metabolism has been linked to cancer development by causing abnormal expression of various genes and dysregulating cytokines and signaling pathways [271]. Most importantly, numerous ASAP1-regulated genes, particularly those that were downregulated when ASAP1 was targeted, displayed significant relevance to relapse-free survival of TNBC patients, indicating ASAP1 as an upstream regulator in driving TNBC progression.

Deconvolution of genetic alterations in cancer genome, such as focal CNAs, has provided an excellent possibility to classify new cancer subtypes and identify novel therapeutic targets for naïve resistant cancers [235, 236, 257]. Our work elucidated extensive amplification-dependent gene expression alterations in TNBC, revealing ASAP1 as a potential TNBC driver functioning upstream of cytokine and apoptosis genes in promoting proliferation and survival. ASAP1 emerges as a potential diagnostic marker as well as therapeutic target for cancer, as ASAP1 has been found to be implicated in multiple oncogenic processes in various cancers [244, 261, 267]. Our results suggest that targeting the upstream regulator ASAP1 and its downstream target genes may provide actionable therapeutic strategies for overcoming the intractable TNBC disease, as well as other resistant cancer types overexpressing ASAP1.

Acknowledgments

JH was financially supported by the China Scholarship Council. We thank Marije Niemeijer for the help with whole transcriptome sequencing analyses.

Funding

This work was supported by the ERC Advanced grant Triple-BC (grant no. 322737).

Authors' contributions

JH, YZ and BvdW conceived and designed the experiments. YZ and BvdW supervised the research. JH, RPM and LvdB performed the experiments. SC performed ADMIRE analyses. LW, JWMM and JAF co-supervised the research. JH, YZ and BvdW wrote the manuscript. All authors read, reviewed and approved the final manuscript.

Additional files: Supplementary data

Additional file 1. Table S1. 148 candidate driver genes for TNBC.

ADAM32	AXIN1	CAMK2B	DBNL	FBXL14	IRF2BP2	MYRFL	PLEKHF1	RASA3	TFDP1
ADAM9	BACE2	CARS	DCUN1D2	FBXW11	KCNMB4	NCOR2	PLXNC1	RBM12B	TGFBR3
ADCY4	BAG4	CCNE1	DDAH1	FGF18	KIAA1217	NPM1	POLD2	RIPK3	TIAL1
ADIPOR2	BAI2	CCT2	DDHD2	FGFR1	KIAA1429	NRIP1	POLM	RRP7A	TM2D2
ADPRHL1	BCL11A	CCT5	DHRS1	FLJ12443	LANCL2	NTRK3	POP4	RTN4IP1	TMEM247
ADSSL1	BNIP1	CDC73	DRD1	GAB1	LCN7	OGG1	PPP1R9A	SACM1L	TMEM67
AIM1	BRD9	CEP170B	DUSP1	GMPR2	LDLRAD4	OLFML2B	PRKWNK1	SF3A1	TMEM8A
ANKH	BRF1	CFH	EGFR	GPR128	LHFP	OTULIN	PROZ	SFXN1	TRIO
ANKRD10	BRPF1	CLPTM1L	EPHA3	GRK5	LMCD1	PANK3	PTK2	SGSM2	TRIP13
ARFGEF3	BTBD6	CNOT2	ERC1	GUCY1A3	LRRC6	PAPOLG	PUS10	SIVA	TTLL3
ARHGEF7	BTG3	CREBRF	F10	HRH2	LTB4R2	PCK2	QRSL1	SPDL1	USP25
ASAP1	C14orf79	CTIF	F7	HTPAP	MET	PDGFRA	RAB20	STK10	VOPP1
ATF6	C19ORF2	CTNNB1	FAM105A	INF2	MRPL28	PEF1	RAD18	TADA3L	WHSC1L1
ATP11A	C6orf203	CUL4A	FAM193A	ING1	MSX2	PHGDH	RAD54B	TARBP1	
ATP4B	CACNA2D	CXADR	FANCD2	IRAK2	MYC	PLEKHA2	RARS	TENS1	

Additional file 3. Table S2. Primary hits of siRNA-mediated loss-of-function screen.

BT549							SUM149PT		
Genes	% control	STDEV	Genes	% control	STDEV		Genes	% control	STDEV
SPDL1	15.78547	0.028709	GMPR2	49.54181	0.119077		SF3A1	20.80776	0.02984
F7	20.16426	0.08556	DDHD2	51.09423	0.045467		TTLL3	28.32151	0.085348
LHFP	22.43687	0.060953	ANKH	51.3599	0.009051		LHFP	29.34977	0.099702
TTLL3	23.45475	0.122612	F10	51.43992	0.059609		CCT2	34.78913	0.101753
SF3A1	26.83166	0.112501	TM2D2	51.55195	0.104935		OGG1	40.60104	0.09454
SIVA	28.4481	0.023476	CCT2	51.94246	0.068448		SIVA	40.79477	0.049073
HRH2	32.91971	0.065549	INF2	52.07369	0.20202		CARS	41.26568	0.011031
FGFR1	34.08162	0.100692	TMEM8A	52.0865	0.049144		SPDL1	45.89733	0.050205
CTIF	34.10723	0.19601	CCT5	53.11398	0.055013		BRF1	46.21027	0.253498
MYC	35.06749	0.00594	LCN7	53.35084	0.007778		EGFR	46.7855	0.001131
CARS	36.17179	0.170483	RRP7A	53.71574	0.07962		HRH2	46.9852	0.221819
OGG1	36.82797	0.00792	RASA3	53.89179	0.017183		BNIP1	49.50071	0.113915
BRF1	37.37852	0.030688	ASAP1	54.25989	0.073115		CCT5	51.82249	0.071701
TMEM247	40.93148	0.05077	ANKRD10	55.46341	0.045538		RRP7A	52.29043	0.135835
POLM	41.42441	0.069155	LTB4R2	55.87632	0.110804		CTIF	52.41859	0.021496
CNOT2	43.43136	0.067104	IRF2BP2	56.17721	0.024112		FBXW11	53.07727	0.056639
DRD1	44.48124	0.074034	BNIP1	57.83205	0.019658		USP25	56.31704	0.055861
TADA3L	44.7085	0.006223	ERC1	58.58106	0.104935		PLEKHF1	56.51971	0.124734
BRPF1	44.81093	0.118087	SFXN1	58.86593	0.039527		BACE2	57.50624	0.108682
CDC73	45.49591	0.123037	TMEM67	59.1252	0.015981		ADPRHL1	58.4898	0.07064
PLEKHF1	49.42338	0.130037							

Additional file 4. Table S3. BC cell line panel.

Cell line	Subtype		Cell line	Subtype		Cell line	Subtype
BT20	TNBC		SUM102PT	TNBC		MDAMB134VI	non_TNBC
BT549	TNBC		SUM1315MO2	TNBC		MDAMB175VII	non_TNBC
DU4475	TNBC		SUM149PT	TNBC		MDAMB330	non_TNBC
HCC1143	TNBC		SUM159PT	TNBC		MDAMB361	non_TNBC
HCC1187	TNBC		SUM185PE	TNBC		MDAMB415	non_TNBC
HCC1395	TNBC		SUM229PE	TNBC		MPE600	non_TNBC
HCC1599	TNBC		SUM52PE	TNBC		OCUBF	non_TNBC
HCC1806	TNBC		BT474	non_TNBC		SKBR3	non_TNBC
HCC1937	TNBC		BT483	non_TNBC		SKBR5	non_TNBC
HCC38	TNBC		CAMA1	non_TNBC		SUM190PT	non_TNBC
HCC70	TNBC		EVSAT	non_TNBC		SUM225CWN	non_TNBC
Hs578T	TNBC		HCC1419	non_TNBC		SUM44PE	non_TNBC
MDAMB157	TNBC		HCC1500	non_TNBC		T47D	non_TNBC
MDAMB231	TNBC		HCC1569	non_TNBC		UACC893	non_TNBC
MDAMB435s	TNBC		HCC1954	non_TNBC		ZR751	non_TNBC
MDAMB436	TNBC		HCC202	non_TNBC		ZR7530	non_TNBC
MDAMB453	TNBC		HCC2218	non_TNBC			
MDAMB468	TNBC		MCF7	non_TNBC			

Additional file 7. Table S4. siASAP1 targeted DEGs in three TNBC cell lines.

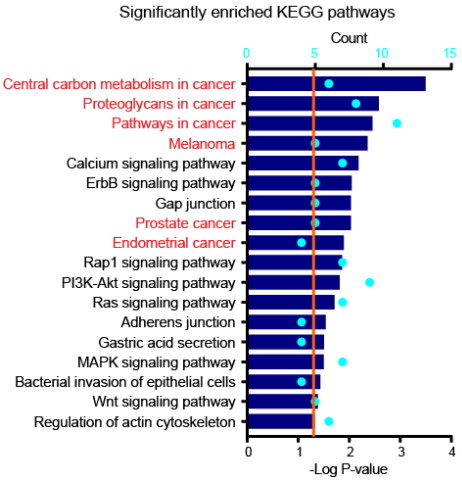
Downregulated genes										Upregulated genes								
17 DEGs pc	18 DEGs pc	16 DEGs pc	44 DEGs pc	95 DEGs pc	133 DEGs p	311 DEGs p	500 DEGs popped up in SUM149			2 DEGs po	14 DEGs pc	22 DEGs pc	41 DEGs pc	79 DEGs pc	117 DEGs p	495 DEGs p	401 DEGs p	
SNCA	RASGRP3	TBC1D5	MGP	**SNCA**	MOCOS	LUM	SNAPIN			PLA2G4C	FNBP4	CNOT4	ARNTL2	**PLA2G4C**	FNBP4	ARNTL2	PHLDB1	
MRPS27	PNMA2	SCG2	MYL6	**MRPS27**	SNCA	HIST3H3	ADORA1			ULK1	TRIM16L	TGFBR1	WDR4	**ULK1**	CASP1	GMEB1	ACADVL	
SREBF2	FHL1	UBL7	ADA	**SREBF2**	TWISTNB	RAPGEF4	TIMELESS				SERPINE1	C1S	SIK1	**FNBP4**	TRIM16L	CHMP7	ATP6V1A	
SSR2	ZNF25	KIF3B	MAGED1	**SSR2**	DKC1	MGP	MELK				SQSTM1	PLPP3	CAB39L	**TRIM16L**	CNOT4	WDR4	TRIM56	
MMGT1	CGREF1	CHST14	ARPIN	**MMGT1**	ENTPD5	SIAE	HIST1H2AE				HMGA2	NFKBIZ	ACADVL	**SERPINE1**	ATXN7L1	KLF6	CMPK1	
BEX3	RBL2	YBX3	TMBIM6	**BEX3**	RASGRP3	MLLT3	PARVB				CD68	C3	E2F3	**SQSTM1**	SLC25A24	CDK5RAP2	UBXN4	
CETN2	RNASE4	CENPH	TXNDC12	**CETN2**	SMC5	NTN5	ADA				CTSD	GBP1	SLC30A1	**HMGA2**	IL31RA	MECP2	IRF6	
PCBD1	C12orf49	HIST1H2A	C1orf198	**PCBD1**	AMMECR1	LRRN1	SH3BP2				AIFM2	SLC7A11	RELL2	**CD68**	SERPINE1	DNAJC11	WWC1	
RPS26	OLFML2B	SLC39A1	TCEAL8	**RPS26**	SV2A	SEMA5B	FKBP11				HELZ2	ARHGAP2	PQLC2	**CTSD**	TGFBR1	RMND5B	E2F3	
ERGIC1	ANKRD13	RELL1	SMUG1	**ERGIC1**	ETV4	MCOLN3	BOC				IRGQ	KIAA1217	ACAP3	**AIFM2**	GBP2	TMEM87A	ZNF248	
UBE2M	SLC38A1	FANCC	PYURF	**UBE2M**	ESPL1	KREMEN1	MTUS1				TAP1	DDX52	DUSP4	**HELZ2**	AKR1B10	SAFB	TGFBR1	
KIAA1191	THOC3	CEP89	TMEM14A	**KIAA1191**	SPPL3	CDON	BEX3				MMP1	EGF	COL16A1	**IRGQ**	TOR1B	GGA1	NFRKB	
TRIQK	CLIC4	FABP5	ATOX1	**TRIQK**	MCRIP2	EPHA3	MAD2L2				KRT15	PANX2	VEGFA	**TAP1**	DDR1	BCAR3	PPP1R21	
PNKD	SLC14A1	LRRC23	EBP	**PNKD**	EARS2	KRTAP5-9	NEMP2				MMP3	EID3	RFTN1	**MMP1**	HSPB8	TVP23C	SLC30A1	
ACAT1	ARL2BP	NME1	HIST1H4E	**ACAT1**	TK1	CFHR1	CHDH					SLC37A2	NRP2	**KRT15**	LANCL2	IGF2R	PPP1R15B	
ANXA1	EMB	IMPDH1	VKORC1	**ANXA1**	ARNT2	ZNF430	C14orf80					FAM71F2	INHBA	**MMP3**	CYTH4	GPX1	BRF2	
DHCR24	RRAGA		PSMB4	**DHCR24**	RACGAP1	RAB20	CDKN2D					TNIP3	ADAM19	**CNOT4**	SQSTM1	ZMYND11	CNOT4	
	EBPL		PDK3	**RASGRP3**	IGFBP3	PIGY	TPRG1					BMF	RNF19B	**TGFBR1**	C1S	BDP1	CSNK1D	
			...									...						
			...									...						
			MCTS1	**ZNF25**	RAB4B	RERG	HIST1H2AM					FGD6	PTPRB	**NFKBIZ**	SENP5	MGAT4B	RHBDF1	

Additional file 8. Table S5. Enrichment of ASAP1-regulated genes in 20 clusters.

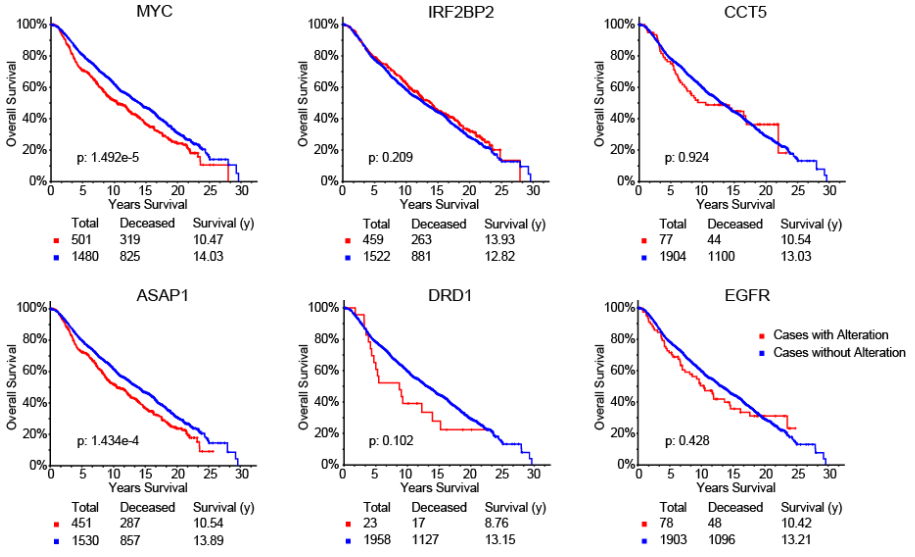
GroupID	Category	Term	Description	LogP	Log(q-valu	InTerm_In	Genes	Symbols
1_Summa	Reactome	R-HSA-128	Cytokine Signaling in Immune system	-7.58181	-3.368	21/690	301,330,18	ANXA1,BIF
2_Summa	GO Biolog	GO:00192	regulation of lipid metabolic process	-6.71756	-3.056	15/397	37,301,476	ACADVL,A
3_Summa	GO Biolog	GO:20012	regulation of apoptotic signaling pathway	-6.67568	-3.056	15/400	90,3383,35	ACVR1,ICA
4_Summa	GO Biolog	GO:00705	response to interleukin-1	-6.67334	-3.056	11/199	301,2633,3	ANXA1,GB
5_Summa	GO Biolog	GO:00109	positive regulation of cell death	-5.99662	-2.673	19/716	301,1871,3	ANXA1,E2I
6_Summa	KEGG Path	hsa04668	TNF signaling pathway	-5.94758	-2.673	8/108	330,2919,2	BIRC3,CXC
7_Summa	GO Biolog	GO:00427	regulation of sulfur metabolic process	-5.48635	-2.449	4/15	5165,6622	PDK3,SNC
8_Summa	GO Biolog	GO:00510	negative regulation of secretion	-5.14609	-2.297	10/232	100,301,19	ADA,ANX
9_Summa	GO Biolog	GO:00507	regulation of inflammatory response	-5.12193	-2.297	14/468	100,301,3	ADA,ANX
10_Summ	KEGG Path	hsa05219	Bladder cancer	-4.96566	-2.170	5/41	1871,1950	E2F3,EGF,(
11_Summ	GO Biolog	GO:00070	cell cycle arrest	-4.93932	-2.167	10/245	1718,3576	DHCR24,C
12_Summ	GO Biolog	GO:00301	extracellular matrix organization	-4.83036	-2.070	12/368	1307,2295	COL16A1,F
13_Summ	GO Biolog	GO:00017	neutrophil apoptotic process	-4.70553	-1.991	3/8	301,3569,2	ANXA1,IL6
14_Summ	GO Biolog	GO:19007	regulation of p38MAPK cascade	-4.62438	-1.931	5/48	3553,7422	IL1B,VEGF,
15_Summ	GO Biolog	GO:00316	response to nutrient levels	-4.24516	-1.648	13/491	38,100,968	ACAT1,AD
16_Summ	Canonical	M145	PID P53 DOWNSTREAM PATHWAY	-4.21964	-1.644	7/137	1509,1871	CTSD,E2F3
17_Summ	Reactome	R-HSA-895	Cellular responses to external stimuli	-4.11338	-1.561	13/506	475,1871,3	ATOX1,E2I
18_Summ	KEGG Path	hsa04010	MAPK signaling pathway	-4.01103	-1.521	9/255	1846,1950	DUSP4,EGI
19_Summ	GO Biolog	GO:00432	regulation of protein complex assembly	-3.92418	-1.474	12/457	3383,4312	ICAM1,MM
20_Summ	GO Biolog	GO:00509	induction of positive chemotaxis	-3.90638	-1.465	3/14	3576,7422	CXCL8,VEC

Additional file 9. Table S6. ASAP1-regulated genes in cytokine signaling, lipid metabolic process and apoptosis signaling pathways.

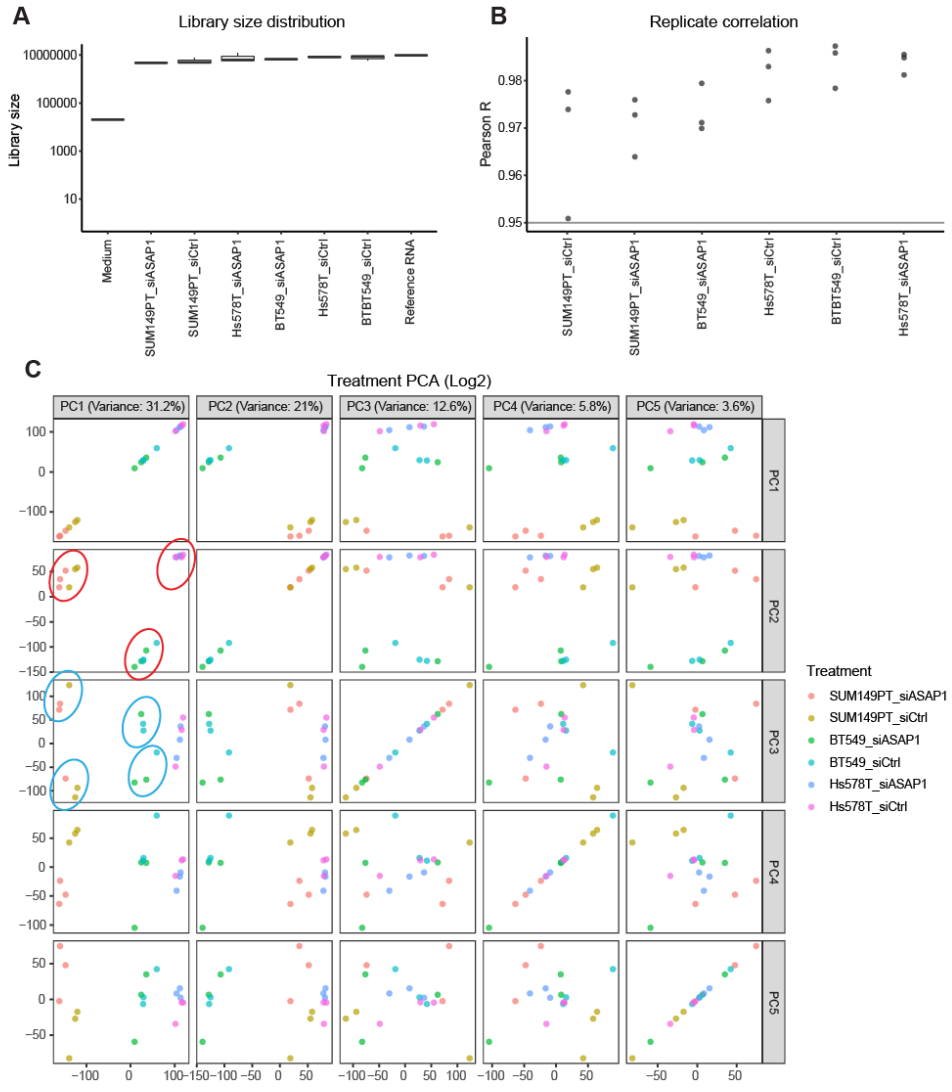
78 cytokine signaling related hits			31 lipid metabolism related hits		33 apoptosis related hits
ACVR1	IRAK2		ACADVL		ACVR1
ADA	KIF3B		ANXA1		ADA
ADAM19	LY6E		ATP1A1		AIFM2
ANAPC5	MAGED1		C3		ANXA1
ANKRD13C	MCM6		FABP5		APEX1
ANXA1	MMP1		IL1B		BEX3
ARL2BP	MMP3		PDK3		BMF
ATP1A1	NFKBIZ		RAN		CD68
BIRC3	NME1		RBL2		COL16A1
BMF	NRP2		SNCA		E2F3
C1S	P3H2		SREBF2		HMGA2
C3	PDK3		PDP2		ICAM1
CABLES1	PLPP3		HELZ2		IL1B
CD68	PNKD		SIK1		IL6
CENPH	PSMB4		WASHC1		INHBA
CETN2	PSMC4		ACAT1		MAGED1
CLIC4	PTPRB		ADA		MAP3K11
CNOT4	RAN		DHCR24		MMP3
CTSD	RBL2		EGF		PNMA2
CXCL1	RFTN1		IMPDH1		SCG2
CXCL2	RPLP0		NME1		SERPINE1
CXCL8	SCG2		PCBD1		SIK1
DUSP10	SERPINE1		EBP		SNCA
DUSP4	SLC7A11		SMUG1		SOS2
E2F3	SNCA		CHST14		SQSTM1
EGF	SOS2		SPP1		SRPX
EMB	SPP1		EBPL		TGFBR1
FHL1	SQSTM1		SLC7A11		TMBIM6
FOXF2	SREBF2		PNKD		TMEM14A
GBP1	TAP1		PLPP3		TRAF1
GBP5	TGFBR1		PYURF		TXNDC12
HIST1H2AI	TMBIM6				UBE2M
HIST1H2BB	TNIP3				YBX3
HIST1H4E	TRAF1				
HMGA2	UBE2M				
ICAM1	VEGFA				
IL1B	WASHC1				
IL6	YBX3				
INHBA	ZGPAT				



Additional file 2. Figure S1. KEGG pathway enrichment analysis of 148 candidate driver genes. Orange line indicated where P-value = 0.05, whereas cyan dots indicated the number of mapped genes in each pathway. Pathways involved in cancer progression were highlighted in red.



Additional file 5. Figure S2. Association of CNA-driven candidate hits, including oncogene MYC and novel driver gene ASAP1, with overall survival of 2173 breast cancer patients. Kaplan Meier plot was generated in cBioPortal using dataset "METABRIC, Nature 2012 & Nat Commun 2016".



Additional file 6. Figure S3. Quality control transcriptomic TemO-Seq analysis. (A) Sequencing library size distribution across treatments. **(B)** Sequencing reproducibility. Pearson correlation coefficient r was calculated among each replicate against triplicate mean per treatment. **(C)** Principle component analysis (PCA) across treatments using Log2 read counts. Red circle, cell line clustering; blue circle, treatment clustering.