TDP 2 modulates the expression of estrogen-responsive oncogenes

Nicholas Manguso[1], Minhyung Kim[1], Neeraj Joshi[1], Rasel Al Mahmud[2, 6], Felipe Cortes-Ledesma[3,8], Xiaojiang Cui[1,4], Shintaro Yamada[2], Junji Ito[2], Shunichi Takeda[2,7], Armando Giuliano[1,4], Sungyong You[1,4,5,] and Hisashi Tanaka[1,4,5]

[1]Department of Surgery, [4]Samuel Oschin Comprehensive Cancer Institute and [5]Biomedical Sciences, Cedars-Sinai Medical Center, West Hollywood, CA, 90048, USA

[2]Department of Radiation Genetics, Graduate School of Medicine, Kyoto University, 606-8501 Kyoto, Japan

[3]Centro Andaluz de Biologia Moleculary Medicina Regenerativa (CAMBIER), CSIC-Universidad de Sevilla (Departmento de genetica) Sevilla, Spain

Corresponding author: Hisashi Tanaka (Hisashi.Tanaka@cshs.org)

Present addresses

[6]Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA

[7]Shenzhen University School of Medicine, Shenzhen, Guangdong, China

[8]Topology and DNA breaks Group, Spanish National Cancer Centre (CNIO), Madrid, Spain

N.M and M.K. contributes equally to this work

Conflict of interest The authors have no conflict of interest to declare regarding this study.

Abstract

Estrogen receptor (ER) is the most prevalent marker of breast cancer. With its ligand estrogen, ER stimulates tumor cell growth by activating a global transcriptional program. The activation involves topoisomerase 2 (TOP2), which resolves topological problems during transcription by transiently creating and re-ligating DNA double-strand breaks (DSBs). Recent studies revealed that other DNA repair proteins also participate in the repair of TOP2-induced DSBs. These noncanonical repair pathways could act as backup processes when TOP2 fails to re-ligate DSBs, but their definite roles remain elusive. Here we examined the role of tyrosyl-DNA phosphodiesterase 2 (TDP2), an enzyme that promotes the removal of abortive TOP2 complex from the 5'-end of the DNA, in the estrogen-induced transcriptome. Using TDP2-deficient ER-positive cells and mice, we showed that TDP2 regulates the expression of oncogene *MYC*. *MYC* induction by estrogen is a very early (1 hour) and TOP2-dependent event. Without TDP2, the induction of *MYC* by estrogen became prolonged and volatile at a later time point. Bulk and single-cell RNA-seq identified oncogenes *MYC* and *CCND1* as genes in which estrogen response was regulated by TDP2. These results could imply that TDP2 plays a role in the repair of TOP2-induced DSBs in specific genomic loci and tightly regulates oncogenic gene expression.

Introduction

With a lifetime risk of one in eight women, breast cancer is the most common female cancer and remains the most frequent cause of cancer-related death in women. About two-thirds of breast tumors express estrogen receptor alpha (ERα), and tumor growth depends on the hormone estrogen [1]. Estrogen plays a crucial role in the development and regulation of female reproductive systems and secondary sex characteristics. 17beta-estradiol (E2) is the predominant

estrogen during reproductive years both in terms of absolute serum levels and estrogen activity. Like all steroid hormones, secreted E2 diffuses into the cell nucleus, where it binds to a receptor ERα [2]. The E2-ERα complex then binds to DNA, acting as potent regulators of gene expression. Transcriptional activation *in vitro* occurs within an hour after E2 stimulation and involves Type II topoisomerase (TOP2), in particular, TOP2 beta (TOP2β) [3, 4].

TOP2 activities release torsional stress during various processes of DNA metabolism, including replication, chromosome segregation, and transcription [5, 6]. TOP2 catalyzes a transient DNA double-stranded break (DSB) into a duplex of DNA and passes another intact duplex through the broken DNA. The catalytic reaction involves a transient intermediate with a covalent bond between the catalytic tyrosine in TOP2 and the 5' phosphate groups of the DNA ends (TOP2 cleavage complex, TOP2cc), followed by the re-ligation of the DSB ends. In general transcription, advancing RNA polymerase II (PolII) on the DNA templates creates over-wound DNA in front and leaves under-wound DNA behind. By relieving torsional stress in front, TOP2 promotes the release of paused PolII for transcriptional elongation [7-9]. Because of the important and broad roles of TOP2 in nucleotide metabolisms, small molecule inhibitors that intervene in each step of TOP2 activity have been developed. TOP2 catalytic inhibitors inhibit the induction of DSBs, and TOP2-poisons stabilize TOP2cc and prevent the re-ligation of TOP2-induced DSBs [10].

In addition to relieving torsional stress in transcription, TOP2 plays a crucial role in the induction of a specific set of genes in response to exogenous stimuli, including E2, heat shock, serum starvation, and neuronal activity [7, 9, 11, 12]. On some of these gene loci, TOP2β was shown to induce DSBs at transcriptional units, such as promoters, which was followed by the recruitment of DSB repair proteins, DNAPK and Ku70 [9, 11]. The recruitment was somewhat

surprising, given that TOP2 is known to be a self-sufficient enzyme, and DSBs induced by TOP2 are transient and re-ligated by TOP2 [5, 13]. TOP2(β) occupies more than 2,000 sites in the genome upon E2 treatment in MCF7 breast cancer cells [14], and a fraction of these sites could possibly recruit DSB repair proteins. In this regard, recent studies revealed noncanonical repair mechanisms of TOP2-induced DSBs that involve two classes of repair proteins; (1) proteins that process (remove) TOP2 that remains covalently attached to the 5' DSB ends (TOP2 cleavage complex, TOP2cc) and (2) those that repair the resulting clean DSB ends. The first process was executed either by 5'-tyrosyl-DNA phosphodiesterase (TDP2) or the Mre11-RAD50-Nbs1 (MRN) complex [15-17]. The resulting clean ends were shown to be repaired by non-homologous end-joining (NHEJ) [18]. BRCA1 would also participate in the repair of DSBs with TOP2cc [17, 19]. Currently, these noncanonical processes are viewed as backup processes; occasionally, TOP2 is trapped at the DNA ends and requires other proteins to remove. It remains elusive whether noncanonical processes are programmed to be a part of the above-mentioned repair of TOP2-induced DSBs in specific gene loci where DSB repair proteins are recruited.

Intriguingly, when left unrepaired, TOP2-induced DSBs can result in transcriptional activation of a specific set of genes. For example, inhibition of the repair of TOP2-induced DSBs by TOP2-poison etoposide resulted in the upregulation of early response genes to neuronal activity in cultured primary neurons [11]. The mechanism underlying the upregulation is of great interest, and it is also crucial to define which genes are upregulated in response to other exogenous stimuli when the repair of TOP2-induced DSBs is inhibited. The upregulation of target genes could cause a serious problem when external stimuli, such as E2 promote cell proliferation. The activation of target genes could potentially lead to abnormal cell proliferation and, ultimately, tumorigenesis. However, it is unknown whether such upregulation can be seen in

a more natural setting of DSB repair deficiency, for example, in cells lacking a DSB repair protein, than etoposide treatment. Etoposide mimics naturally stalled TOP2cc at the 5' end of DSBs. Stalled TOP2cc by etoposide occurs very frequently in the entire cell population, in which transcriptional upregulation can be manifested and detected by techniques for measuring the average gene expression level in a cell population, such as qPCR, digital PCR, and bulk RNA-seq. In contrast, in cells with a deficiency of a DSB repair protein, stalled TOP2cc may not occur efficiently in the entire cell population. In such a case, the techniques described above may not be able to detect transcriptional upregulation. However, upregulation can be seen in DSB repair-deficient cells if TOP2-induced DSB is programmed to be repaired by the DSB repair protein in a certain gene locus. Such a possibility was shown at the protein level in mice lacking NHEJ repair; E2 stimulated the expression of MYC protein in mammary epithelial cells of mice lacking DNAPKcs [20].

We wished to identify a set of E2-induced genes that becomes upregulated in cells deficient in TOP2-induced DSB repair. The list of such genes would also inform a potentially novel process; gene loci where noncanonical repair of TOP2-induced DSBs are programmed. To do so, we created ER-positive breast cancer cell lines that lack TDP2, and investigated a global transcriptional program. We chose TDP2 because TDP2 knockout cells are viable and amenable to E2 stimulation experiments. Using digital PCR, bulk- and single cell-RNAseq, we identified that *MYC* was among several genes that exhibited enhanced transcription in response to E2 in TDP2-KO breast tumor cells relative to the TDP2-proficient cells. We also found that another breast cancer oncogene, *CCND1*, exhibited enhanced transcription when TDP2 was deficient. These results suggest that noncanonical repair processes of TOP2-induced DSBs would tightly

regulate oncogenic gene expression. The results may also explain why deficiency of DSB repair proteins often underlies the susceptibility of breast tumors.

Results

MYC induction by 17beta-estradiol (E2) depends on TOP2

MYC is a well-studied oncogene and transcriptionally responds to E2 [21]. We extracted RNA from two ER-positive breast cancer cell lines MCF7 and T47D, before and after treatment with E2 (Figure 1A). We used a digital PCR assay to measure the expression of *MYC* mRNA. In this assay, a *MYC* Taqman probe labeled with a dye FAM and a control probe, either *TBP* or *TFRC*, labeled with a dye VIC were mixed with cDNA from cell lines and PCR-amplified in ~20,000 x 1 nL reaction (Fig. 1B). The *MYC* expression levels (the number of *MYC*-positive 1 nL reactions) were normalized by the internal control (the number of internal control-positive reactions). We added 10 nM E2 and incubated cells for two hours before extracting RNA. Fold induction of *MYC* was determined by (normalized *MYC* expression with E2 treatment / normalized MYC expression without E2 treatment).

ER-positive (ER+) cell lines, MCF7 and T47D, increased *MYC* expression after E2 treatment (2.59-fold and 2.17-fold, respectively). In contrast, the ER-negative (ER-) cell line MDA-MB-231 did not show the induction (0.93-fold change) (Figure 1C). Further, we sought to understand the temporal regulation of *MYC* expression, and measured *MYC* mRNA at 1, 2, and 6 hours after the addition of 10 nM E2. In both the MCF7 and T47D cell lines, *MYC* expression was highest at 1 hour (MCF7 2.89-fold, T47D 2.79-fold) followed by gradual decreases until 6 hours (Figure 1D). A previous study determined that TOP2 mediates the induction of E2-responsive genes [3]. Therefore, we next tested whether *MYC* induction is also TOP2-dependent. We used a TOP2

catalytic inhibitor Merbarone which prevents TOP2 from creating DSBs [22]. We added Merbarone to cell culture media 1 hour before adding 10 nM E2, and incubated cells for 2 hours. We observed a dose-dependent inhibition of *MYC* induction in both MCF7 and T47D (Figure 1E), strongly suggesting that E2 stimulation of *MYC* expression in ER-positive cells is TOP2-dependent.

MYC transcriptional activation in TDP2 knockout cells

Since Merbarone inhibits DSB induction, DSBs could play a crucial role in the transcriptional activation of *MYC*. When left unrepaired, TOP2-induced DSBs activate transcription for a specific set of genes [11]. Therefore, we next sought to determine whether unrepaired DSBs could further enhance the *MYC* induction by E2. To test the idea, we knocked out *TDP2* in ER-positive cell lines. When TOP2 fails to re-ligate DSBs and is aborted, proteasome degrades the TOP2. TDP2 removes the remnant aborted TOP2, and creates a clean DSB for NHEJ machinery to repair [18]. We used CRISPR-Cas9 to knockout TDP2 in both MCF7 (MCF7 TDP2KO-1) and T47D cell lines. In addition, we employed an independently created TDP2 knockout MCF7 cell line (MCF7 TDP2KO-2)[17]. We confirmed the lack of TDP2 protein in knockout cells by western blotting (Figure 2A). We also confirmed the TDP2 deficiency functionally using the TOP2 poison, etoposide. Etoposide traps TOP2cc at DSBs, which would require TDP2 to repair. Both the MCF7 and T47D cells deficient in TDP2 were significantly more sensitive to etoposide than parental cell lines (Figure 2B).

We followed the dynamics of *MYC* induction using digital PCR as in Figure 1C. We noticed that the highest level of induction at 1 hour after the addition of E2 (10nM) was similar between MCF7 and MCF7 TDP2KO-1 cells (Figure 2C). However, the *MYC* expression continued to be high in MCF7 TDP2KO cells, whereas it declined quickly in MCF7 cells. We

further measured the induction of *MYC* at 2 hours after the addition of 10 nM E2 and found the increased fold induction of *MYC* in all three TDP2 knockout cell lines relative to their parental TDP2 wild-type counterparts (Figure 2D). Because the expression was measured by total RNA of entire cell populations, these results suggest that *MYC* induction by E2 was enhanced in a significant proportion of the TDP2KO cell population. A prevailing view is that alternative repair proteins, including TDP2, only repair abortive TOP2-induced DSBs that could occur in a very small subset of cells. However, these results suggest that a significant fraction of cells requires TDP2 to suppress *MYC* induction. The delayed repair of TOP2-induced DSBs could promote the *MYC* response to E2 at 2 hours, as previously described for early response genes to neuronal activity [11]. Consistently, we observed that etoposide promotes the induction of *MYC* in MCF7 cells (Fig. 2E, left). Inhibiting NHEJ could also promote the *MYC* response, as previously shown *in vivo* in DNA-PKcs deficient mice [20]. To test this possibility, we inhibited NHEJ by DNA-PKcs inhibitor NU7441 [23]. Indeed, *MYC* induction at 2 hours after the addition of E2 was increased by NU7441 (5 μM) (Fig. 2E, right).

Induction of MYC by estrogen *in vivo* in mice mammary glands

To address the relevance of the above observation *in vivo*, we examined *MYC* transcriptional response to E2 in mice mammary glands. We injected E2 (300 μg/kg body weight) or PBS (control) intraperitoneally into three *TDP2* wild-type and *TDP2*$^{-/-}$ mice [18]. Mice before injection (0 hour) and 6 and 20 hours after injection were examined for MYC protein expression in CK8/18-positive mammary epithelial cells by immunostaining (Fig. 3A). We counted MYC-positive epithelial cells in 3 to 12 mammary ducts at each timepoint and scored the fraction of MYC-positive cells in each mammary duct (Fig. 3B). MYC-positive cells increased in both wild-type and *TDP2*$^{-/-}$ mice at 6 hours. MYC-positive cells become less

frequent at 20 hours, showing transient induction of MYC in mammary epithelial cells. MYC-positive cells in *TDP2*$^{-/-}$ mice at 20 hours were significantly more frequent than in *TDP* wild-type mice ($p<0.01$) at 20 hours or in *TDP2*$^{-/-}$ mice without E2 at 0 hour ($p<0.01$). The prolonged induction of MYC in *TDP2*$^{-/-}$ mice is reminiscent of TDP2KO MCF7 cells (Fig. 2C) and is consistent with our observation *in vitro*.

MYC and MYC downstream targets induced by E2

MYC is a beta Helix-Loop-Helix transcription factor that, with a partner protein Max, binds to DNA at CACGTG E-Box sequence and modulates global gene expression [24]. To examine the expression of MYC-regulated genes in MCF7 and MCF7 TDP2KO cells, we conducted RNA-seq analysis using total RNA before and 2 hours after the addition of 10nM E2 (bulk RNA-seq). We found that a large number of genes become differentially expressed in each MCF7 (2,021) and MCF7 TDP2 KO cells (1,034) (Supplementary Table 1 and 2; fold change ≥ 1.5 and combined P-value < 0.05), and *MYC* was one of them in both cell types. Consistent with the results from digital PCR, *MYC* was upregulated more in MCF7 TDP2KO cells (3.16 fold) than parental MCF7 cells (1.94 fold) by E2 (Fig. 4A), confirming the prolonged response of *MYC* to estrogen when TDP2 is deficient.

To assess whether *MYC* is activated after E2 treatment in MCF7 TDP2KO and parental MCF7 cells, we performed Gene Set Enrichment Analysis (GSEA) [25] with predefined MYC target genes. GSEA determines whether a defined set of genes shows statistically significant, concordant differences between two biological states. We looked for the significant changes in expression of genes whose promoters have MYC binding sites within the 4 kb of the transcription start sites (MYC_Q2), and therefore potentially direct targets of MYC between E2-treated and untreated cells (Fig. S1, A, top). When treated with E2 for 2 hours, the vast majority

of genes in MYC_Q2 were nonrandomly distributed and primarily found to be upregulated in both MCF7 and MCF7 TDP2KO cells. We further calculated the normalized enrichment score (NES) (Fig. S1, B). NES reports the degree of overrepresentation of genes at the extremes of the entire ranked genes. We found that MYC_Q2 genes were slightly more clustered in highly upregulated genes in MCF7 TDP2KO cells than in MCF7 cells. We also examined gene sets, the expression of which were increased by overexpressing MYC in human primary mammary epithelial cells (Hallmark MYC target v2 and MYC_UP.V1_UP) [26]. Genes in both sets were enriched in upregulated genes, with equivalent NES values between MCF7 and MCF7 TDP2KO. These results suggest that the upregulation of MYC by E2 modulated global transcriptional programs in both cell types, especially when TDP2 was deficient.

MYC induction at the single-cell level

Bulk RNA-seq provides us with the average RNA expression of the entire cell population. The observed upregulation of *MYC* in MCF7 TDP2KO cells could occur either (1) when only a subset of cells upregulates *MYC* at an extremely high level (bimodal distribution) and, as a result, increase the average of the population, or (2) when *MYC* is induced more robustly in the vast majority of MCF7 TDP2KO cell population than that of MCF7 cells (normal distribution). The first scenario would indicate that, for those DSBs that are involved in the regulation of *MYC* after E2 treatments, TOP2-induced DSBs are primarily re-ligated by TOP2 itself, and TDP2 only participates in the repair in a subset of cells. The second scenario would imply that TDP2 is programmed in the repair of TOP2-induced DSBs that regulate *MYC* expression. These possibilities can be distinguished by single-cell RNA-seq (sc RNA-seq) experiments.

Four groups of cells were processed for scRNA-seq: MCF7 parental cells, MCF7 parental cells treated with E2 for 2 hours, MCF7 TDP2KO cells, and MCF7 TDP2KO cells treated with

E2 for 2 hours. The quality control and normalization of scRNA-seq data were conducted using the Seurat pipeline [27]. After filtering out poor quality cells and genes, the transcriptome data included 354 genes in MCF7 with E2, 585 genes in MCF7 without E2, 708 genes in TDP2KO with E2, and 399 genes in TDP2KO without E2. With the filtered and normalized data, we performed two independent comparisons: E2-treated MCF7 versus untreated MCF7 and E2-treated TDP2KO versus untreated TDP2KO. With adjusted p-value <0.05 and log fold change ≥ 0.1 (1.11 fold), we identified a total of 111 genes that were differentially expressed between untreated and E2-treated MCF7; 100 upregulated and 11 down-regulated genes, respectively (Supplementary Table 3). E2 treatment in TDP2-KO led to gene expression changes in 327 genes: 138 up- and 189 down-regulated genes (Supplementary Table 4).

Consistent with the digital PCR and bulk RNA-seq, *MYC* expression was induced by E2 treatment in MCF7 TDP2KO and MCF7 (log2-fold change 0.43 (1.54 fold) and 0.17 (1.19 fold), Fig. 4A). And, the *MYC* induction by E2 was higher in MCF7 TDP2KO cells than in MCF7. Normalized expression data showed that MCF7 TDP2KO cell populations exhibited continuous distributions without distinct subpopulations and uniquely contained cells with very high expression of *MYC* (normalized expression >1.5) after E2 treatment (Fig. 4B). Histograms of raw read count were consistent with the skewed unimodal distribution in all four populations (Fig. 4C). The normal distribution of expression values of single cells suggested that the vast majority of cells in the population contributed equally to the mean increase, which supports scenario (2).

We next examined whether the enhanced transcriptional response of the *MYC* is associated with the differences of cellular processes between MCF7 and MCF7 TDP2KO cells, such as cell cycle phases. Cell cycle phases are considered confounders in single-cell studies [28]. To address this issue, we allocated individual cells to G1, S, and G2/M phases using a single

transcriptome data-based cell cycle prediction tool [29]. We then tested whether cell cycle distribution confounds for the observed difference in *MYC* transcriptional responses by collecting *MYC* expression profiles with and without E2 in each cell cycle phase (Fig. 4D). Based on the cell cycle phase prediction, we found that MCF7 TDP2KO cells are modestly more enriched with G1 cell populations than MCF7 cells; >60% of TDP2 KO cells were in G1, while 50% of MCF7 cells were in G1 (Fig. S2). However, the *MYC* transcriptional induction by E2 was less pronounced between MCF7 and MCF7 TDP2KO cells in G1 cells (1.26-fold with $p<0.0001$ in MCF7 vs. 1.47-fold with $p<0.0001$ in MCF7 TDP2KO, Wilcoxon rank-sum test) than in S (1.08-fold with $p=0.0332$ vs. 1.52-fold with $p<0.0001$) and G2M cells (1.24-fold with $p=0.0014$ vs. 1.88-fold with $p<0.0001$) (Fig. 4C). Therefore, the difference in the distribution of cell cycles was not a major contributor to increased *MYC* induction in MCF7 TDP2KO cells. Instead, cell cycle profiling revealed that the increased response to E2 would likely come from cells in S/G2M phases.

The dominance of upregulated genes by the short period of E2 treatment

The global transcriptional responses to E2 have been extensively studied in MCF7 cell line and described both upregulated and downregulated genes. However, the amounts of E2 and the durations of the treatments were unique to each experiment. It was also reported that the repertoire of E2-responsive genes differs between genomic platforms (e.g., microarray, RNA-seq and scRNA-seq) [30, 31]. We consider that the intersections between different genomic platforms (bulk RNA-seq and scRNA-seq) would provide confident sets of E2 (10nM)-responsive genes at an early time point (2 hr).

Bulk RNA-seq identified a large number of genes (2021 for MCF7 and 1034 for MCF7 TDP2KO) that are differentially expressed between untreated and E2-treated cells. The numbers

of differentially expressed genes (DEGs) in our scRNA-seq data were smaller: 111 genes for MCF7 and 327 genes for MCF7 TDP2KO (Fig. 5A). Thirty-two genes were common between bulk and single-cell approaches in MCF7, while 62 genes were found to be common in MCF7 TDP2KO cells. The majority of common, DEGs were upregulated genes in both cell lines (Fig. 5B; Supplementary Table 5 and 6). In MCF7, 28 genes were induced and 4 genes were repressed, with one gene GPRC5A in opposite directions between bulk- and sc-RNA-seq. In MCF7 TDP2KO, 58 genes were induced and 4 genes were repressed. Among these genes, there were 8 common DEGs between MCF7 and MCF7 TDP2KO; *MYC*, *CCND1*, *NRIP*, *SH3BP5*, *IGFBP4*, *RET*, *NBPF1* and *CA12* (Fig, 5C). Notably, *MYC* and *CCND1* are oncogenes in breast cancer, the roles of which in tumorigenesis are well established [32-35]. These results suggest that estrogen-responsive genes at an early time point (2h) are predominantly upregulated genes, and TDP2 could regulate responses to E2 in both *MYC* and *CCND1*.

Discussion

In this study, we evaluated the mechanisms underlying the transcriptional responses to 17beta-estradiol (E2), a dominant estrogen, in ER-positive breast cancer cells and breast epithelial cells in mice. Transcriptional response to estrogen is known to involve Topoisomerase 2 (TOP2) [3], which creates transient DNA double-strand breaks, relieves torsional stress and reseal the ends. Recent studies showed that a set of proteins involved in DNA metabolism, including TDP2, Mre11, BRCA1 and NHEJ proteins also play a role in TOP2-induced DSBs [15-18]. We wished to obtain insights on how these proteins contribute to the repair of TOP2-induced DSBs. To do so, we monitored transcriptional response to E2 in TDP2 wild-type and TDP2KO MCF7 and T47D cell line. Defects in the repair of TOP2-induced DSB were shown to upregulate transcriptional response to external stimuli [11]. Therefore, we anticipated that the

degree of changes in E2 response would be proportional to the fraction of the population experiencing repair defects due to the lack of TDP2.

We confirmed that transcription of an oncogene *MYC* responded to E2 (10 nM) very early, as early as 1 hour after the addition of E2 (Fig. 1D). Early response was abrogated when TOP2-induced DSB was inhibited by Merbarone, a catalytic inhibitor of TOP2, in a dose-dependent manner in both MCF7 and T47D cell lines (Fig. 1D), strongly suggesting that TOP2 is involved in the early transcriptional response of *MYC*. The transcriptional response is most likely through TOP2-induced DSBs, as TOP2 poison etoposide enhanced the transcriptional response (Fig. 2E). This notion was further supported by the TDP2 knockout experiments in vitro with various platforms for gene expression measurements; *MYC* transcriptional response was upregulated in two cell lines, two KO clones of MCF7 and *in vivo* (Fig. 2, 3 and 4). Because the enhanced *MYC* response was also seen in cells treated with an NHEJ inhibitor, unrepaired TOP2-induced DSBs were likely the cause of upregulation (Fig. 2E). Importantly, we were able to detect the upregulation by the techniques that measured the average expression level in the entire cell population, such as digital PCR and bulk mRNA-seq. Therefore, TDP2 seems to be involved in the transcriptional regulation of *MYC* in a significant fraction of the cell population.

Using single-cell RNA-seq, we further found that the observed enhanced response of *MYC* in MCF7 TDP2KO cells relative to MCF7 cells was driven by the population-wide upregulation of *MYC* by E2. This was evidenced by the unimodal distribution of *MYC* mRNA in MCF7 TDP2KO cells (Fig. 4C). Unimodality of gene expression was seen for *MYC* and other commonly upregulated genes. In a previous study of scRNA-seq for the E2-stimulated transcriptional response in MCF7, a large proportion of genes showed a bimodal distribution of gene expression [36]. However, the numbers of cells captured for transcriptomic analysis are a

smaller in the study (87 MCF7 cells) relative to our study (354 cells from MCF7 with E2, and 585 cells from MCF7 without E2) and may confound the difference from our study. With the results along with the regulation of *MYC* expression response to E2 by TOP2 (Fig. 1E), we speculate that TDP2-dependent repair of TOP2-induced DSBs is programmed for regulating *MYC* expression when stimulated by E2. Unrepaired DSBs could suppress the transcription of nearby genes [37-39]. On the contrary, the accumulation of DNA damages at stimuli-induced genes was associated with transcriptional upregulation by releasing PolII from pausing [11]. Also, unrepaired DSBs far from the target genes, such as enhancer elements, would not directly disrupt the target genes and elongate mRNA. This scenario may be the case for *MYC*, in which enhancer elements are known to be dispersed within a few Mb gene deserts surrounding *MYC* [40, 41]. In particular, an enhancer located at 135kb distal to *MYC* (Myc+135kb enhancer) has an ER binding site and responds to estrogen treatments [42]. The estrogen-induced expression of enhancer RNA and the physical interaction with *MYC* promoter was greatly promoted when the repair of TOP2-induced DSBs was inhibited by an ATM inhibitor (KU-55933). Although ATM and TDP2 function independently to repair etoposide-induced DSBs [43], and these two repair pathways' spatial and temporal regulations remain unclear, TDP2 would also regulate the estrogen-induced activation of *MYC* enhancers through the repair of TOP2-induced DSBs.

scRNA-seq also allowed us to investigate the responses to E2 in each cell cycle phase. We found that TDP2 KO cells responded to E2 more robustly in S and G2/M phases than in G1 phase. There are two pathways for the removal of abortive TOP2ccs from the ends: TDP2 and Mre11/BRCA1. It is currently unknown how these two pathways separate the tasks for the removal. BRCA1 supports the nuclease activity of Mre11 to remove abortive TOP2ccs [17]. This activity was shown to be active in G1 cells. With the robust upregulation of *MYC* in S and G2/M

cells in our study, we speculate that these two pathways repair DSbs with abortive TOP2cc in different cell cycle phases. A recent study showed that TDP2 binds to K63 and K27 poly-ubiquitinated cellular proteins and this binding stimulates TDP2-catalyzed removal of TOP2 from DNA [44]. The identity of the K27 and/or K63 poly-ubiquitinated proteins bound to TDP2 is not known, but such a protein could regulate TDP2 activity in a cell-cycle dependent manner.

We employed bulk and scRNA-seq for evaluating the expression of E2-responsive genes at 2-hour time point. TDP2-proficient MCF7 cell line has been used to study E2-responsive genes [30, 31]. These studies differ in their experimental conditions, including measuring platforms, the amounts of E2, and the duration of treatments. As a result, these studies showed large variations in the numbers and spectrums of E2-responsive genes. Genetic variations between MCF7 sublines [45] can also contribute to the variable results. In this regard, we used the same clones and identical experimental conditions for two RNA-seq platforms to minimize the noises associated with the experimental design. The genes that were common between the two platforms were strongly restricted by scDNA-seq data in both MCF7 and TDP2-KO cells, as only small fractions of differentially expressed genes detected by bulk RNA-seq were also called by scRNA-seq. This is likely because scRNA-seq data are in general very heterogeneous with many zero values [46]. Nonetheless, these platform-independent, differentially expressed genes showed a striking propensity towards upregulated genes, including oncogenes *MYC* and *CCND1* that are frequently upregulated in ER-positive breast tumors. Estrogen induces mammary ductal hyperplasia through MYC in mice deficient in NHEJ, which is downstream of TDP2-mediated repair of TOP2cc. CCND1 is more frequently upregulated in ER-positive tumors than other subtypes and promotes cell cycle progression [47, 48]. By participating in the regulation of these oncogenes, TDP2 would be a candidate suppressor of the most common subtype of breast tumors.

Materials and Methods

Cell culture and digital PCR MCF7 and MCF7 TDP2KO cells were maintained with MEM with 10% FBS. T47D, T47D TDP2KO, and MDA-MB-231 cells were cultured with DMEM with 10% FBS. TDP2KO cells were created by targeting exon 2 of the *TDP2* gene (5'-GGCTCAGAGATGGTTTCAGGT-3'). The targeting vector was constructed using pSpCas9(BB)-2A-GFP(PX458), a gift from Feng Zhang (Addgene plasmid #48138).

Etoposide sensitivity was determined using the colony formation assay. Briefly, 3000 cells were plated in each well of a 6-well plate, and cells were cultured for 8 days with or without etoposide before staining with crystal violet. Experiments were performed in triplicate. Colonies were counted using the Hybrid Cell Function of the Keyence BZ-X700.

RNA was extracted using the RNeasy mini kit, and first-strand cDNA synthesis was performed using Superscript III (Invitrogen). Cells were cultured in phenol red-free medium supplemented with 5% FBS-charcoal stripped for 24 hours then treated with 10 nM 17β-estradiol for either 1, 2 or 6 hours before extracting RNA. NU7441 was added to the media 3 hours before adding 10 nM E2. Etoposide was added to the media 1 hour before adding E2. Digital PCR was done using the QuantStudio 3D Digital PCR System. A TaqMan probe labeled with FAM (MYC-Hs00153408, IL20-Hs00218888 or OLFML3-Hs01113293 and VIC (TBP-Hs00427620 or TFRC-Hs00951083) were mixed in a reaction, and the amount of transcript (copies/microliter) was determined using the Analysis Suite (Thermo Fisher Scientific).

Mouse experiments

TDP2 Wild-type and *TDP2*$^{-/-}$ strains came from the C57BL6/J strain. E2 (300 μg/kg body weight) was administered via intraperitoneal injection (IP) into three independent two-month-old female mice at 100 μL volume. As a control, the same volume of an E2 solvent (PBS) is injected. Mice were sacrificed to harvest mammary gland tissue, and frozen blocks were prepared. After a brief wash of the isolated tissues with cold PBS, tissues were fixed with paraformaldehyde (4%, Cat# 163-20145, Wako, Japan) for 15 min at 4° C and then washed 3 times with 1x PBS. To preserve the tissue integrity, incubation was done in 30% sucrose (in PBS) for 30 min to 3 hours. Tissues were then embedded into optimal cutting temperature (OCT) compound (Cat# 4583, Sakura, Japan) into Cryomolds (Cat# 4565, Sakura, Japan). For cryo-sectioning of frozen blocks, a cryostat (CM1850, Leica, Germany) was used at -25° C and sectioned into 10 μm slices. Then slides were heated at 55° C for 10 to 30 min to dry. Dried slides were washed three times with Tween-20 in PBS (0.05 %, PBS-T). For blocking 1 % BSA and 5 %, goat serum in PBS-T was used for 3 hours at room temperature, then washed once with PBS-T. Slides were then incubated with both α-Cytokeratin-8/18 (1/10, rat monoclonal, University of Iowa, US) and α-c-Myc antibody (1/200 [Y69], ab32072, rabbit monoclonal, Abcam) overnight at 4° C. After washing with PBS-T, slides were incubated with both α-rabbit (Alexa Fluor 546) and α-rat (Alexa Fluor 488) secondary antibodies (Molecular probe, US). c-Myc and Cytokeratin-8/18 signals were detected using LEICA SP8.

Data preprocessing and normalization of bulk RNA-seq data

MCF7, MCF7 with 10 nM E2 treatment for 2 hours, MCF7 TDP2KO, MCF7 TDP2KO with 10 nM E2 treatment for 2 hours were subject to RNA-seq at the Cedars-Sinai Genomics Core. Biological duplicates were taken for all of the processes. Our quality control pipeline utilizes the Illumina standards to assess base call quality and truncates low-quality reads. The quality of

sequencing data was assessed using MultiQC software [49]. Sequence alignment and quantification were performed using the STAR-RSEM pipeline [50, 51]. Reads overlapping exons in the annotation of Genome Reference Consortium Human Build 38 (GRCh38) will be identified. Genes were filtered out and excluded from downstream analysis if they failed to achieve raw read counts of at least 2 across all the libraries. Batch effects were corrected by ComBat [52]. The trimmed mean of M-values normalization method (TMM) [53] (version 1.6.1) was used for calculating normalized count data.

Differential expression analysis of bulk RNA-seq data

For the two comparisons, 1) E2 treated versus non-treated in MCF7 cells and 2) E2 treated versus non-treated in MCF7 TDP2KO cells, we used the integrated hypothesis testing method as previously reported [54]. Briefly, for each gene, two P-values were computed by performing two-tailed Student's T-test and two-tailed median different test using the empirical distributions that were estimated by random permutations of the samples. We then combined two P-values into a combined P-value using Stouffer's method. Multiple testing correction was done by using Storey's correction method [55]. Finally, the DEGs were selected with the ones having a combined p-value < 0.05 and fold-change ≥ 1.5. The GSEA was used to check the enrichment of gene sets [25].

Single-cell RNA-seq data preprocessing and differential expression analysis

Single-cell RNA-seq was performed using the 10X Genomics scRNA-seq approach at the Cedars-Sinai Genomics Core. The Cell Ranger Single-Cell Software Suite was used for demultiplexing, barcode assignment, and unique molecular identifiers (UMIs) quantification. The reads were aligned to the hg38 reference genome using a pre-built annotation package obtained from the 10X Genomics website. All lanes per sample were processed using the

'cellranger count' function. The output from different lanes was then aggregated using 'cellranger aggr' with –normalise set to 'none.' The quality control and analysis were conducted using the Seurat packages in R (version 3.2.2) [27]. Poor-quality cells were identified by filtering out the cells with the number of detected genes less than 2,000. In addition, all cells with 11.89% (three standard deviations) or more of UMIs mapping to mitochondrial genes were defined as non-viable or apoptotic and removed from the analysis. Genes detected in at least three cells were used for further analysis, resulting in a total of 19,043 genes remaining for the analysis. We also ensured that none of the reads in our data set were derived from index swapping. For this, we labeled cells with barcodes that appeared in more than one sample (non-unique barcodes) as 'NULL' and excluded these cells for further analysis. Gene expression values were normalized by log normalization using the "NormalizeData" function in Seurat. We performed two comparisons of 1) E2 treated versus non-treated in MCF7 cells and 2) E2 treated versus non-treated in MCF7 TDP2KO cells. The "FindMarkers" function in Seurat was used with filtered data to identify DEGs between the cell groups. DEGs were selected with adjusted p-value <0.05 and log fold change ≥ 0.1 (1.11 fold). Fold change threshold was determined at 95 percentile value of the empirical null distribution of fold changes, which was generated by 1,000 times random permutation of the cells.

Statistical Analysis

We performed principal component analysis (PCA) for visualization of the samples to assess the distribution of the samples by gene expression profile. We used MATLAB (v.9.0; Mathworks, Natick, MA, USA), R (v.3.5), and Python (v.3.7) for bioinformatic analysis.

Acknowledgment

We thank Cedars-Sinai Cancer, Applied Genomics, Computation and Translational Core and Flow Core for technical support. This work is supported by Cedars-Sinai CTSI Clinical Scholar Award (to N.M.), NIHR01CA149385, Cedars-Sinai Cancer Biology Development Fund and Institutional Support (to H.T.), Clinical Scholar Award (to N.M), Leukemia & Lymphoma Society (to N.J.) and Margie and Robert E. Petersen Foundation, The Fashion Footwear Charitable Foundation of New York, Inc., Avon Foundation and Associates for Breast and Prostate Cancer Studies (to A.G.).

Data availability

Bulk and single-cell RNA-seq data are available in Gene Expression Omnibus (accession # GSE225423).

Author Contributions

Conceptualization, N.M., M.K., S.T., S.Y.Y., and H.T.; Data collection, N.M., N.J., R.A.M., and H.T.; Data analysis and interpretation, N.M., M.K., R.A.M., S.Y.Y., and H.T.; Materials, F.C-L., S.Y., J.I., X.C.,; Writing the manuscript, N.M., M.K., S.T., S.Y.Y., and H.T.; Funding Acquisition, N.M., A.E.G., and H.T.

References

1 Katzenellenbogen BS, Katzenellenbogen JA. Estrogen receptor transcription and transactivation: Estrogen receptor alpha and estrogen receptor beta: regulation by selective estrogen receptor modulators and importance in breast cancer. *Breast Cancer Res* 2000; 2: 335-344.

2 Fuentes N, Silveyra P. Estrogen receptor signaling mechanisms. *Adv Protein Chem Struct Biol* 2019; 116: 135-170.

3 Ju BG, Lunyak VV, Perissi V, Garcia-Bassets I, Rose DW, Glass CK *et al*. A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription. *Science* 2006; 312: 1798-1802.

4 Hah N, Danko CG, Core L, Waterfall JJ, Siepel A, Lis JT *et al*. A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell* 2011; 145: 622-634.

5 Pommier Y, Sun Y, Huang SN, Nitiss JL. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat Rev Mol Cell Biol* 2016; 17: 703-721.

6 Morimoto S, Tsuda M, Bunch H, Sasanuma H, Austin C, Takeda S. Type II DNA Topoisomerases Cause Spontaneous Double-Strand Breaks in Genomic DNA. *Genes (Basel)* 2019; 10.

7 Tubbs A, Nussenzweig A. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell* 2017; 168: 644-656.

8 King IF, Yandava CN, Mabb AM, Hsiao JS, Huang H-S, Pearson BL *et al*. Topoisomerases facilitate transcription of long genes linked to autism. *Nature* 2013; 501: 58-62.

9 Bunch H, Lawney BP, Lin Y-F, Asaithamby A, Murshid A, Wang YE *et al*. Transcriptional elongation requires DNA break-induced signalling. *Nature Communications* 2015; 6: 10191.

10 Nitiss JL. Targeting DNA topoisomerase II in cancer chemotherapy. *Nat Rev Cancer* 2009; 9: 338-350.

11 Madabhushi R, Gao F, Pfenning AR, Pan L, Yamakawa S, Seo J *et al*. Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. *Cell* 2015; 161: 1592-1605.

12 Calderwood SK. A critical role for topoisomerase IIb and DNA double strand breaks in transcription. *Transcription* 2016; 7: 75-83.

13 Nitiss JL. DNA topoisomerase II and its growing repertoire of biological functions. *Nature Reviews Cancer* 2009; 9: 327-337.

14 Manville CM, Smith K, Sondka Z, Rance H, Cockell S, Cowell IG *et al*. Genome-wide ChIP-seq analysis of human TOP2B occupancy in MCF7 breast cancer epithelial cells. *Biol Open* 2015; 4: 1436-1447.

15 Ledesma FC, El Khamisy SF, Zuma MC, Osborn K, Caldecott KW. A human 5′-tyrosyl DNA phosphodiesterase that repairs topoisomerase-mediated DNA damage. *Nature* 2009; 461: 674-678.

16 Hoa Nguyen N, Shimizu T, Zhou Zhong W, Wang Z-Q, Deshpande Rajashree A, Paull Tanya T *et al*. Mre11 Is Essential for the Removal of Lethal Topoisomerase 2 Covalent Cleavage Complexes. *Molecular Cell* 2016; 64: 580-592.

17 Sasanuma H, Tsuda M, Morimoto S, Saha LK, Rahman MM, Kiyooka Y *et al*. BRCA1 ensures genome integrity by eliminating estrogen-induced pathological topoisomerase II–DNA complexes. *Proceedings of the National Academy of Sciences* 2018; 115: E10642-E10651.

18 Gómez-Herreros F, Romero-Granados R, Zeng Z, Alvarez-Quilón A, Quintero C, Ju L *et al*. TDP2-dependent non-homologous end-joining protects against topoisomerase II-induced DNA breaks and genome instability in cells and in vivo. *PLoS Genet* 2013; 9: e1003226.

19 Aparicio T, Baer R, Gottesman M, Gautier J. MRN, CtIP, and BRCA1 mediate repair of topoisomerase II–DNA adducts. *Journal of Cell Biology* 2016; 212: 399-408.

20 Itou J, Takahashi R, Sasanuma H, Tsuda M, Morimoto S, Matsumoto Y *et al*. Estrogen Induces Mammary Ductal Dysplasia via the Upregulation of Myc Expression in a DNA-Repair-Deficient Condition. *iScience* 2020; 23: 100821-100821.

21 Shang Y, Hu X, DiRenzo J, Lazar MA, Brown M. Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* 2000; 103: 843-852.

22 Drake FH, Hofmann GA, Mong SM, Bartus JO, Hertzberg RP, Johnson RK *et al*. In vitro and intracellular inhibition of topoisomerase II by the antitumor agent merbarone. *Cancer Res* 1989; 49: 2578-2583.

23 Zhao Y, Thomas HD, Batey MA, Cowell IG, Richardson CJ, Griffin RJ *et al*. Preclinical evaluation of a potent novel DNA-dependent protein kinase inhibitor NU7441. *Cancer Res* 2006; 66: 5354-5362.

24 Grandori C, Cowley SM, James LP, Eisenman RN. The Myc/Max/Mad network and the transcriptional control of cell behavior. *Annu Rev Cell Dev Biol* 2000; 16: 653-699.

25 Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA *et al*. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* 2005; 102: 15545-15550.

26 Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D *et al*. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 2006; 439: 353-357.

27 Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd *et al*. Comprehensive Integration of Single-Cell Data. *Cell* 2019; 177: 1888-1902 e1821.

28 Hsiao CJ, Tung P, Blischak JD, Burnett JE, Barr KA, Dey KK *et al*. Characterizing and inferring quantitative cell cycle phase in single-cell RNA-seq data analysis. *Genome Res* 2020; 30: 611-621.

29 Nestorowa S, Hamey FK, Pijuan Sala B, Diamanti E, Shepherd M, Laurenti E *et al*. A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* 2016; 128: e20-31.

30 Cheung E, Kraus WL. Genomic Analyses of Hormone Signaling and Gene Regulation. *Annual Review of Physiology* 2010; 72: 191-218.

31 Jagannathan V, Robinson-Rechavi M. Meta-analysis of estrogen response in MCF-7 distinguishes early target genes involved in signaling and cell proliferation from later target genes involved in cell cycle and DNA repair. *BMC Syst Biol* 2011; 5: 138.

32 Escot C, Theillet C, Lidereau R, Spyratos F, Champeme MH, Gest J *et al*. Genetic alteration of the c-myc protooncogene (MYC) in human primary breast carcinomas. *Proc Natl Acad Sci U S A* 1986; 83: 4834-4838.

33 Sicinski P, Donaher JL, Parker SB, Li T, Fazeli A, Gardner H *et al*. Cyclin D1 provides a link between development and oncogenesis in the retina and breast. *Cell* 1995; 82: 621-630.

34 Weinstat-Saslow D, Merino MJ, Manrow RE, Lawrence JA, Bluth RF, Wittenbel KD *et al*. Overexpression of cyclin D mRNA distinguishes invasive and in situ breast carcinomas from non-malignant lesions. *Nat Med* 1995; 1: 1257-1260.

35 D'Cruz CM, Gunther EJ, Boxer RB, Hartman JL, Sintasath L, Moody SE *et al*. c-MYC induces mammary tumorigenesis by means of a preferred pathway involving spontaneous Kras2 mutations. *Nature Medicine* 2001; 7: 235-239.

36 Zhu D, Zhao Z, Cui G, Chang S, Hu L, See YX *et al*. Single-Cell Transcriptome Analysis Reveals Estrogen Signaling Coordinately Augments One-Carbon, Polyamine, and Purine Synthesis in Breast Cancer. *Cell Rep* 2018; 25: 2285-2298 e2284.

37 Harding SM, Boiarsky JA, Greenberg RA. ATM Dependent Silencing Links Nucleolar Chromatin Reorganization to DNA Damage Recognition. *Cell Rep* 2015; 13: 251-259.

38 Larsen DH, Hari F, Clapperton JA, Gwerder M, Gutsche K, Altmeyer M *et al*. The NBS1-Treacle complex controls ribosomal RNA transcription in response to DNA damage. *Nat Cell Biol* 2014; 16: 792-803.

39 Shanbhag NM, Rafalska-Metcalf IU, Balane-Bolivar C, Janicki SM, Greenberg RA. ATM-dependent chromatin changes silence transcription in cis to DNA double-strand breaks. *Cell* 2010; 141: 970-981.

40 Bahr C, von Paleske L, Uslu VV, Remeseiro S, Takayama N, Ng SW *et al*. A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. *Nature* 2018; 553: 515-520.

41 Ahmadiyeh N, Pomerantz MM, Grisanzio C, Herman P, Jia L, Almendro V *et al*. 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. *Proc Natl Acad Sci U S A* 2010; 107: 9742-9746.

42 Najnin RA, Al Mahmud MR, Rahman MM, Takeda S, Sasanuma H, Tanaka H *et al*. ATM suppresses c-Myc overexpression in the mammary epithelium in response to estrogen. *Cell Reports* 2023; 42: 111909.

43 Álvarez-Quilón A, Serrano-Benítez A, Ariel Lieberman J, Quintero C, Sánchez-Gutiérrez D, Escudero LM *et al*. ATM specifically mediates repair of double-strand breaks with blocked DNA ends. *Nature Communications* 2014; 5: 3347.

44 Schellenberg MJ, Appel CD, Riccio AA, Butler LR, Krahn JM, Liebermann JA *et al*. Ubiquitin stimulated reversal of topoisomerase 2 DNA-protein crosslinks by TDP2. *Nucleic Acids Res* 2020; 48: 6310-6325.

45 Ben-David U, Siranosian B, Ha G, Tang H, Oren Y, Hinohara K *et al*. Genetic and transcriptional evolution alters cancer cell line drug response. *Nature* 2018; 560: 325-330.

46 Risso D, Perraudeau F, Gribkova S, Dudoit S, Vert J-P. A general and flexible method for signal extraction from single-cell RNA-seq data. *Nature Communications* 2018; 9: 284.

47 Roy PG, Pratt N, Purdie CA, Baker L, Ashfield A, Quinlan P *et al*. High CCND1 amplification identifies a group of poor prognosis women with estrogen receptor positive breast cancer. *International Journal of Cancer* 2010; 127: 355-360.

48 Ortiz AB, Garcia D, Vicente Y, Palka M, Bellas C, Martin P. Prognostic significance of cyclin D1 protein expression and gene amplification in invasive breast carcinoma. *PLOS ONE* 2017; 12: e0188068.

49 Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011; 12: 323.

50 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2012; 29: 15-21.

51 Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016; 32: 3047-3048.

52 Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2006; 8: 118-127.

53 Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* 2010; 11: R25.

54 You S, Yoo S-A, Choi S, Kim J-Y, Park S-J, Ji JD *et al*. Identification of key regulators for the migration and invasion of rheumatoid synoviocytes through a systems approach. *Proceedings of the National Academy of Sciences* 2014; 111: 550-555.

55 Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* 2003; 100: 9440-9445.

Figure Legends

Figure 1 Estrogen and TOP2-dependent induction of *MYC*

(A) Experiment design. MCF7 and T47D cells were treated with 10 nM 17β-estradiol for 1, 2, and 6 hours. RNA was extracted at each time point for 3D PCR analysis and RNA-seq at 2 hours.

(B) 3D PCR output from Analysis Suites. Blue dots represent wells only with the amplification of genes detected by FAM (*MYC*), red dots represent wells only with the amplification detected by VIC (*TBP* or *TFRC*) probe (internal control), yellow dots represent wells without amplification, green dots represent wells with the amplification of both probes. Normalized *MYC* expression was calculated by the ratio of FAM (*MYC*) to VIC (*TBP* or *TFRC*).

(C) *MYC* induction in cells after 2 hours of treatment with 10 nM E2 relative to cells with no E2 in two ER-positive breast cancer cell lines (MCF7 and T47D) and an ER-negative cell line (MD-MBA-231).

(D) Dynamic responses of *MYC* induction. Cells were treated with 10 nM E2 for 0, 1, 2, and 6 hours. *MYC* fold-induction peaked at 1 hour and remained elevated until 6 hours.

(E) A TOP2 catalytic inhibitor Merbarone abrogated the induction of *MYC* mRNA. MCF7 and T47D cells were treated with increasing concentrations of Merbarone for 1 hour then treated with 10 nM E2 for 2 hours. *MYC* induction was progressively decreased in a concentration-dependent manner in both cell lines. * - $p<0.05$, **- $p<0.01$

Figure 2 TDP2 suppresses the volatile and prolonged induction of *MYC* by estrogen

(A) Western blot showing MCF7-1, T47D and MCF7-2 cell lines and TDP2 KO cells for each cell line.

(B) Etoposide sensitivity measured by colony formation assay for MCF7 (top) and T47D (bottom). Results from three independent experiments. * - $p<0.05$, **- $p<0.01$

(C) *MYC* transcriptional responses persist in TDP2-deficient cells. MCF7 and MCF7 TDP2 KO cells were treated with 10 nM E2 for 1, 2, and 6 hours. After a similar level of induction at 1

hour, high level of response persisted in *MCF7* TDP2 KO cells persisted through 6-hour time point. Results from two independent experiments. * - $p<0.05$, **- $p<0.01$

(D) *MYC* transcriptional responses at 2 hours. MCF7-1, MCF7-2 and T47D, and their TDP2 KO counterparts were treated with 10 nm E2 for 2 hours. Note that two independent TDP2 KO clones of MCF7 cells were tested. Results from three independent experiments.

(E) NU7441 and etoposide increased the *MYC* transcriptional response. MCF7 cells, cultured with or without wither NU7441 or etoposide for three hours were then treated with 10 nM E2 for 2 hours. Results from three independent experiments.

Figure 3 The induction of MYC protein by E2 *in vivo*

(A) Representative images of mammary ducts from wild-type (left) and TDP2KO (right) mice immunostained with MYC (red) and CK-8/18 (green) proteins before injection (0 hours) and 6 and 20 hours after injection of E2.

(B) BOX plots showing the fraction of MYC-positive cells in (CK8/18-positive) mammary epithelial cells. Each plot represents the distribution of the fraction of MYC-positive epithelial cells in a mammary duct. MYC-positive epithelial cells were scored in 3 to 12 mammary ducts in each mouse at each timepoint.

Figure 4 Bulk and Single-cell RNA sequencing analysis of *MYC* transcriptional response by E2

(A) *MYC* transcriptional response to E2 (2 hours) in MCF7 and MCF7 TDP2KO cells, measured by bulk RNA-seq and single-cell RNA-seq.

(B) Violin plots show normalized MYC expression distributions in MCF7 and MCF7 TDP2KO cells with or without E2 (2 hours), obtained from single-cell RNA-seq data.

(C) Histograms showing the skewed unimodal distribution of MYC expression in MCF7 and MCF7 TDP2KO cells with or without E2 (2 hours), obtained from single-cell RNA-seq data.

(D) Violin plots showing the normalized expression of *MYC* at the single-cell level in G1, S, and G2/M cell populations.

Figure 5 Commonly upregulated genes between bulk- and single-cell RNA-seq

(A) Venn diagrams showing the differentially expressed genes by E2 treatment in MCF7 and MCF7 TDP2KO cells. 32 genes in MCF7 and 62 genes in MCF7 TDP2KO are common between the two platforms. The significances of overlap were calculated using Fisher's exact test.

(B) Bar plots showing the number of common DEGs by E2 (2 hours). 28 and 4 genes were commonly up- and down-regulated genes in E2 treated MCF7 versus untreated MCF7, respectively. 58 and 4 genes were commonly up- and downregulated in E2 treated MCF7 TDP2KO versus untreated MCF7 TDP2KO, respectively.

(C) Venn-diagram of 32 and 62 common DEGs from bulk and single cell RNA-seq data and table of 8 overlaping genes. The significance of overlap was calculated using Fisher's exact test.

Figure 1 Manguso et al,

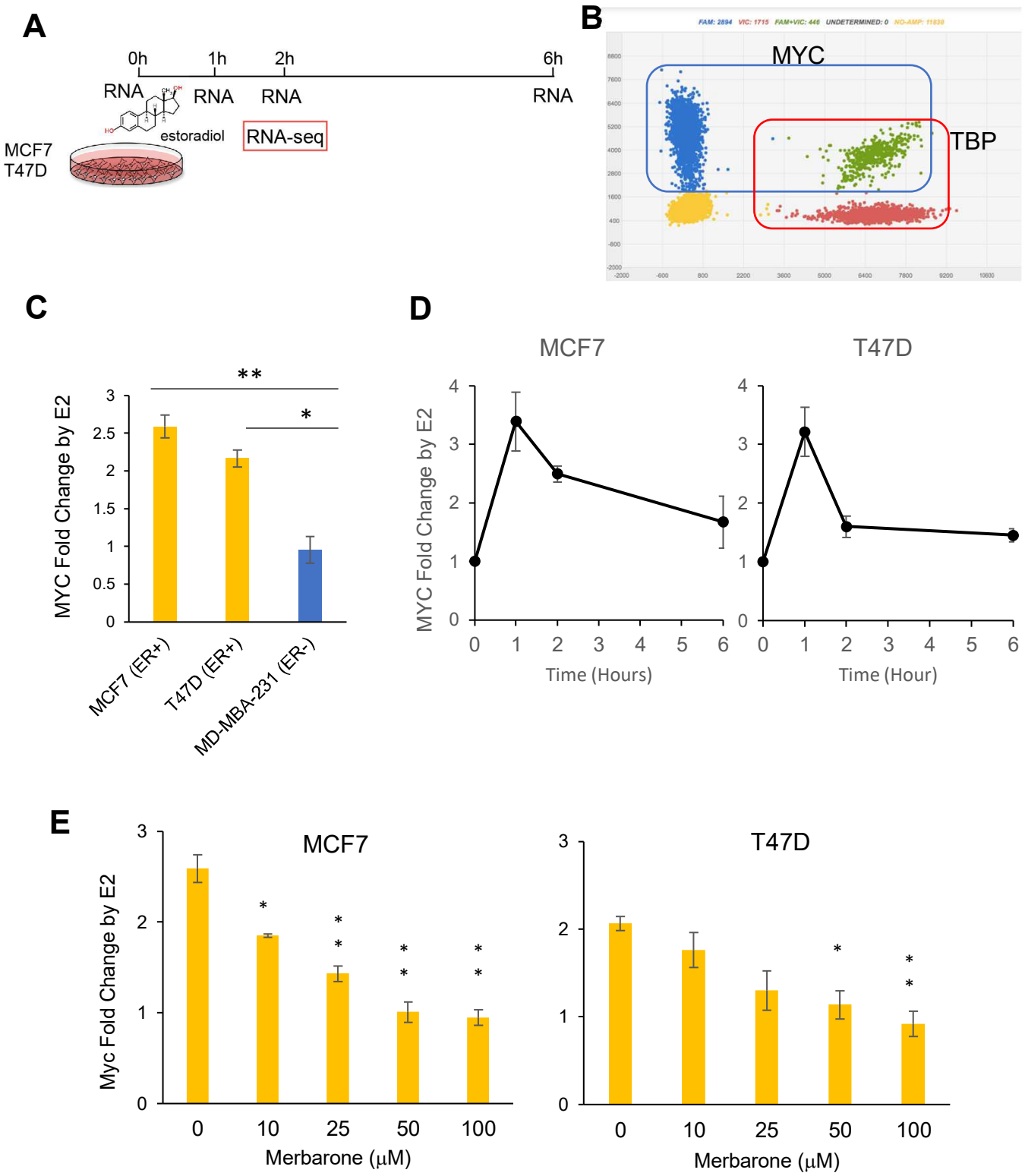


Figure 2 Manguso et al,

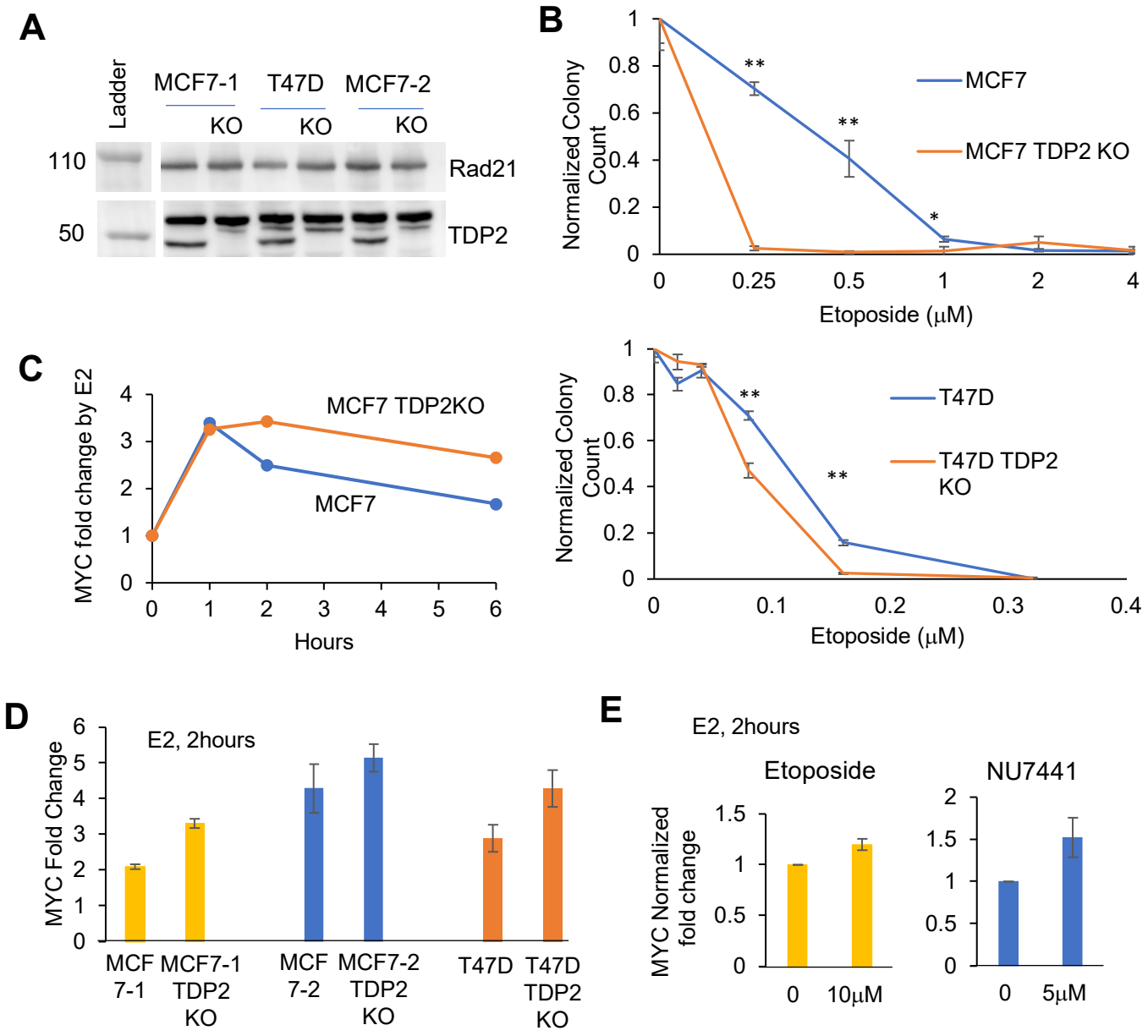


Figure 3, Manguso et al,

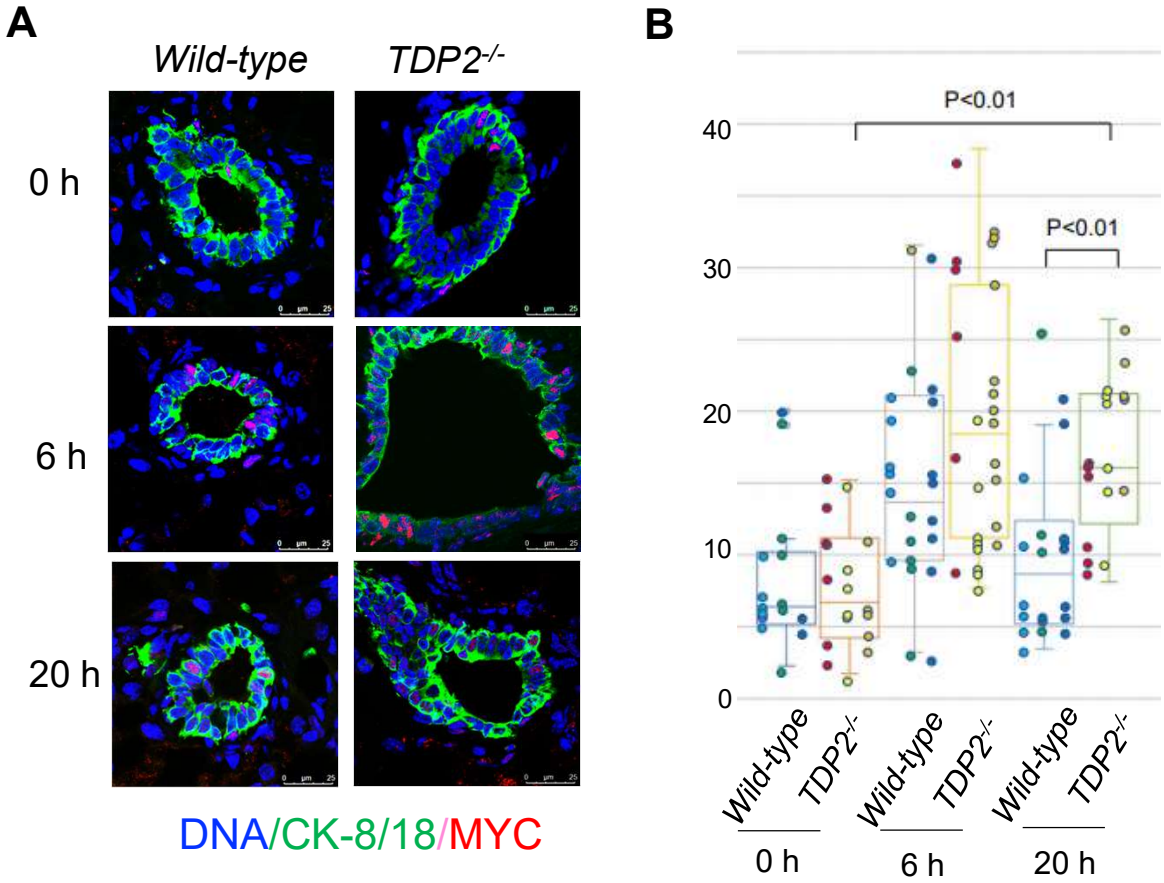
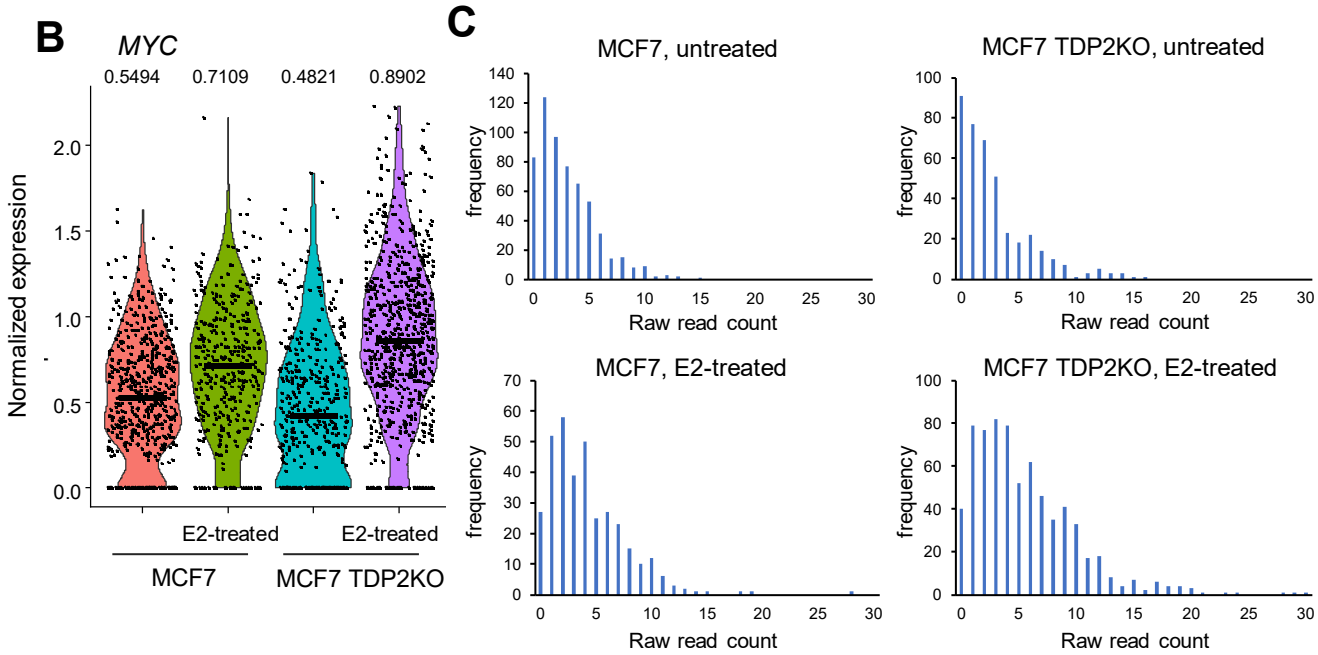


Figure 4, Manguso et al,

A

Comparison	Bulk RNA-seq log2-FC	Bulk RNA-seq P-value	Single-cell RNA-seq log-FC	Single-cell RNA-seq P-value
MCF7 with E2 vs without E2	0.9760	0.0194	0.1722	<0.0001
MCF7 TDP2KO with E2 vs without E2	1.6595	0.0001	0.4267	<0.0001



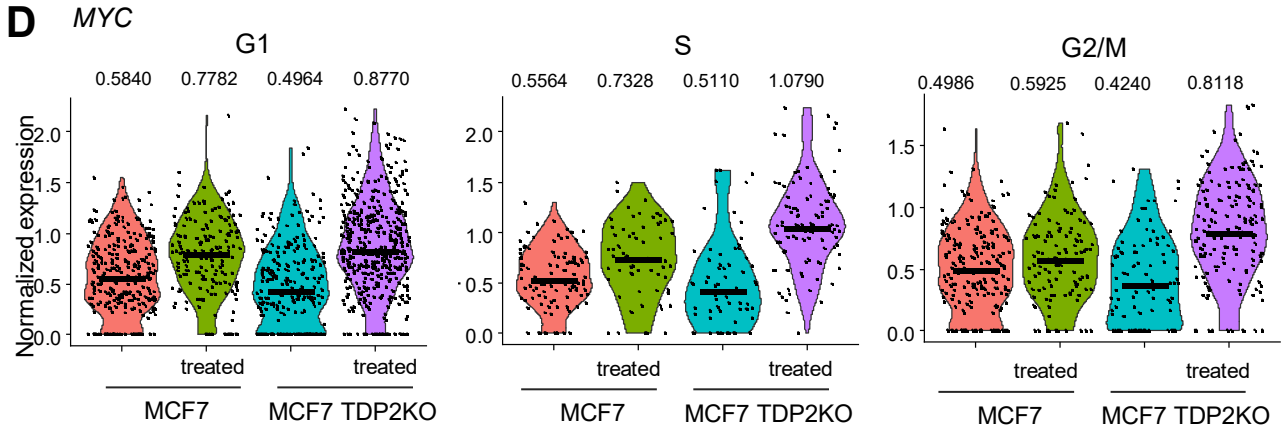
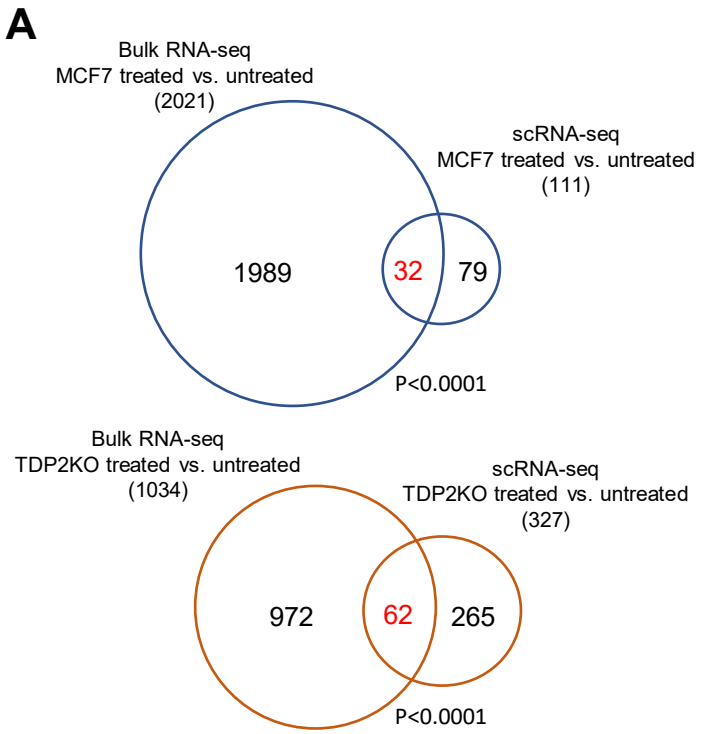
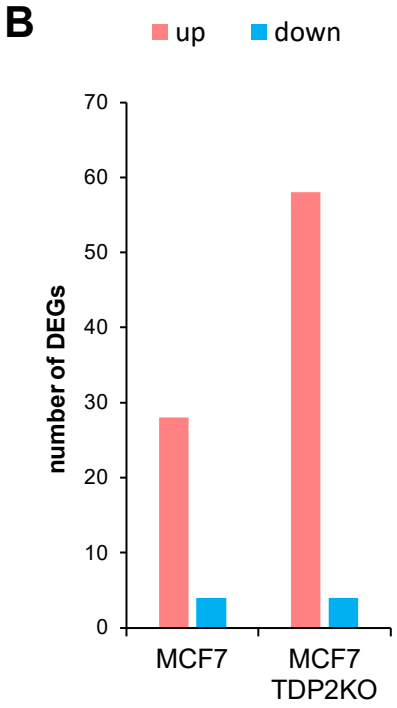
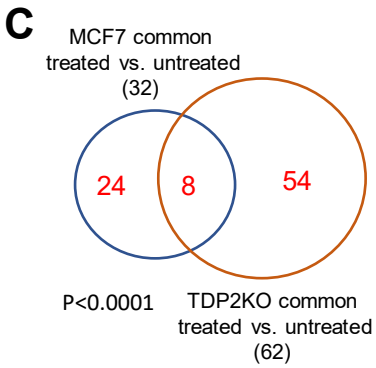


Figure 5, Manguso et al,







	Bulk RNA-seq				scRNA-seq			
	MCF7		MCF7 TDP2KO		MCF7		MCF7 TDP2KO	
	log FC	p value	log FC	p value	log FC	p value	log FC	p value
MYC	0.98	0.02	1.66	<0.01	0.17	3.8E-06	0.43	1.3E-41
CCND1	0.61	0.01	0.81	0.01	0.27	4E-15	0.35	1.2E-35
NRIP1	1.61	<0.01	1.97	<0.01	0.15	9E-06	0.19	2.7E-24
SH3BP5	0.64	0.01	0.61	0.01	0.14	0.01	0.25	5.6E-24
IGFBP4	1.27	<0.01	0.81	0.04	0.14	<0.01	0.39	3.5E-39
RET	1.45	0.01	0.99	0.03	0.18	5.4E-16	0.16	2.4E-13
NBPF1	0.86	0.02	0.90	0.02	0.15	3E-07	0.14	4.1E-08
CA12	1.21	0.01	0.62	0.05	0.24	2.3E-15	0.58	2.9E-77