

 Open access • Posted Content • DOI:10.1101/2021.02.19.431930

An mRNA expression-based signature for oncogene-induced replication-stress

— [Source link](#) 

Sergi Guerrero Llobet, Arkajyoti Bhattacharya, Marieke Everts, Bert van der Vegt ...+2 more authors

Institutions: University Medical Center Groningen

Published on: 19 Feb 2021 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Gene, Cancer, Oncogene, Gene expression and KRAS

Related papers:

- [A core MYC gene expression signature is prominent in basal-like breast cancer but only partially overlaps the core serum response.](#)
- [Genomic and Proteomic Analysis Reveals a Threshold Level of MYC Required for Tumor Maintenance](#)
- [Combined microarray analysis of small cell lung cancer reveals altered apoptotic balance and distinct expression signatures of MYC family gene amplification](#)
- [Genetic Alterations and Oncogenic Pathways Associated with Breast Cancer Subtypes](#)
- [Delineating Genetic Alterations for Tumor Progression in the MCF10A Series of Breast Cancer Cell Lines](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/an-mrna-expression-based-signature-for-oncogene-induced-g5bkw5wzea>

Title:

An mRNA expression-based signature for oncogene-induced replication-stress

Authors:

Sergi Guerrero Llobet MSc^{1,3} Arkajyoti Bhattacharya MSc^{1,3}, Marieke Everts BSc¹, Bert van der Vegt MD-PhD², Rudolf S.N. Fehrmann MD-PhD^{1,4}, Marcel A.T.M. van Vugt PhD^{1,4,5}

Affiliation:

1. Department of Medical Oncology, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713GZ, Groningen, the Netherlands
2. Department of Pathology, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713GZ, Groningen, the Netherlands
3. shared first author
4. shared senior authorship
5. correspondence; Marcel van Vugt, PhD, m.vugt@umcg.nl

ABSTRACT

Background

Oncogene-induced replication stress characterizes many aggressive cancers, including triple-negative breast cancer (TNBC). Several drugs are being developed that target replication stress, although it is unclear how tumors with high levels of replication stress can be identified. We aimed to develop a gene expression signature of oncogene-induced replication stress.

Methods

TNBC and non-transformed RPE1-*TP53*^{wt} and RPE1-*TP53*^{mut} cell lines were engineered to overexpress the oncogenes *CDC25A*, *CCNE1* or *MYC*. DNA fiber analysis was used to measure replication kinetics. Analysis of RNA sequencing data of cell lines and patient-derived tumor samples (TCGA n=10,592) was used to identify differential gene expression. Immunohistochemical validation was conducted on breast cancer samples (n=330).

Results

RNA sequencing revealed 52 commonly upregulated genes after induction of *CDC25A*, *CCNE1* or *MYC* in our cell line panel. Integration with gene expression data of TCGA samples with amplification of replication stress-inducing oncogenes (*CDC25A*, *CCNE1*, *MYC*, *CCND1*, *MYB*, *MOS*, *KRAS*, *ERBB2*, and *E2F1*), yielded a six-gene signature of oncogene-induced replication stress (*NAT10*, *DDX27*, *ZNF48*, *C8ORF33*, *MOCS3*, and *MPP6*). Expression of *NAT10* in breast cancer samples was correlated with phospho-RPA ($R=0.451$, $p=1.82 \times 10^{-20}$) and γ H2AX ($R=0.304$, $p=2.95 \times 10^{-9}$). Applying the oncogene-induced replication stress signature to patient samples (TCGA n=8,862 and GEO n=13,912) defined the replication stress landscape across 27 tumor subtypes, and identified diffuse large B cell lymphoma, ovarian cancer, TNBC and colorectal carcinoma as cancer subtypes with high levels of oncogene-induced replication stress.

Conclusion

We developed a gene expression signature of oncogene-induced replication stress, which may facilitate patient selection for agents that target replication stress.

INTRODUCTION

A subset of cancers, including triple-negative breast cancer (TNBC) shares a profound genomic instability^{1,2} This phenomenon is characterized by continuous gains and losses of chromosomal fragments and complex genomic rearrangements. As a consequence, genomic instability underlies the rapid acquisition of genomic aberrations that drive therapy failure³. Finding new treatment options for genomically unstable cancers is not only relevant for TNBC, but also other hard-to-treat cancers with extensive genomic instability, including high-grade serous ovarian cancer (HGSOC)⁴. Genomic instability is observed in multiple aggressive cancer subtypes and is associated with the inability of cancer cells to faithfully repair DNA damage⁵. Increasingly, an important source of the DNA lesions that fuels genomic instability appears to be replication stress^{6,7}.

During S-phase of the cell cycle, all DNA must be replicated in a coordinated manner, which is initiated at genomic loci called 'replication origins'⁸. Replication origins are fired in a temporally-controlled way, which prevents exhaustion of the nucleotide pool and warrants the availability of essential components of the replication machinery⁹. DNA replication can be challenged in various ways, which is collectively referred to as replication stress. An important cause of replication stress is the uncoordinated initiation of origin firing due to oncogene activation^{7,10,11}. Consequently, oncogene activation leads to depletion of the nucleotide pool and collisions between of the replisome and the transcription machinery, resulting in slowing or complete stalling of replication forks^{7,9,12}. When stalled replication forks are not resolved in time, they can collapse and cause DNA double-strand breaks (DSBs) and lead to genomic instability.

Several oncogenes have been linked to induction of replication stress, many of them leading to aberrant activation of CDK2. For instance, overexpression of the CDK2-binding partner Cyclin E1 (*CCNE1*) or the CDK2-activating phosphatase CDC25A leads to replication slow-down and reversal of replication forks¹³. In line with this notion, amplification of *CCNE1* has been proposed as a biomarker for tumors with high levels of replication stress¹⁴. However, multiple other oncogenic events beyond *CCNE1* amplification also can lead to replication stress, including overexpression of *MYC*^{15,16}, *MOS*¹⁷, *E2F1*¹¹, or expression of the E6/E7 HPV oncoproteins⁹. Currently, there is no uniform way to determine oncogene-induced replication stress levels in cancer samples.

Identification of cancers with high levels of replication stress is increasingly relevant because several drugs have been developed that specifically target tumor cells with high levels

of replication stress. Inhibition of WEE1 has been shown to have therapeutic efficacy in HGSOC, which is considered as a prototypical tumor type with high levels of replication stress¹⁸. Interestingly, patients with the largest decrease in tumor size upon WEE1 inhibitor treatment showed enrichment for *CCNE1* amplification¹⁸. In line with this notion, overexpression of *CCNE1* sensitized TNBC cell lines to WEE1 inhibition^{14,19}. Furthermore, an unbiased genomic screen identified regulators of CDK2 as determinants of WEE1 inhibitor sensitivity²⁰. Mechanistically, WEE1 inhibition is thought to exacerbate levels of replication stress further, while it inactivates the G2/M cell cycle checkpoint, driving cells into mitotic catastrophe²¹. Similarly, inhibition of the ATR or CHK1 checkpoint kinases has been shown to preferentially target tumor cells with molecular characteristics that are associated with replication stress²²⁻²⁴.

It is currently unclear how patients can be optimally selected for treatment with agents that target replication stress. To this end, we performed gene expression profiling of cell lines with oncogene-induced replication. Further refinement with expression profiles of patient derived tumor samples yielded a gene expression signature of replication stress, which allowed us to describe a pan-cancer landscape of oncogene-induced replication stress.

RESULTS

Overexpression of CDC25A, CCNE1, or MYC results in replication stress

To develop an mRNA expression-based signature for oncogene-induced replication stress, we engineered a panel of cell lines to overexpress CDC25A, CCNE1, or MYC in a doxycycline-inducible manner. This cell line panel included non-transformed human retina epithelial (RPE1) cell lines, either with wild type *TP53* (RPE1-*TP53*^{wt}) or a derivative in which *TP53* was mutated using CRISPR-Cas9 (RPE1-*TP53*^{mut}), and TNBC cell lines MDA-MB-231, BT549 and HCC-1806 (Fig. 1A, Suppl. Fig. 1). To validate that oncogene induction indeed affected DNA replication, DNA fiber analysis was performed (Fig. 1B). A severe reduction in DNA synthesis velocity was observed upon induction of CDC25A, CCNE1, or MYC, as measured by IdU fiber tract lengths in RPE1-*TP53*^{wt} cells (32%, 23% and 42% decrease in CDC25A, CCNE1, and MYC overexpressing cells versus controls, respectively, Fig. 1C). Similar decreases in DNA replication dynamics were observed in RPE1-*TP53*^{mut} cells (34%, 27%, and 51% decrease in CDC25A, CCNE1 and MYC overexpressing cells versus controls, respectively, Fig. 1D), indicating that these effects are independent of *TP53* status. Subsequently, the impact of oncogene expression on DNA synthesis was analyzed in TNBC cell lines. Again, we consistently observed shortening of IdU tract lengths in MDA-MB-231, BT549 and HCC-1806 cell lines upon doxycycline-induced overexpression of CDC25A, CCNE1 or MYC (Fig. 1E-G), but not in empty vector controls (Fig. 1C-G). Taken together, these data indicate that induction of CDC25A, CCNE1, or MYC results in replication stress in cancer cells, as well as in untransformed cell lines, independently of *TP53* status.

Gene expression profiling of cell line models with oncogene-induced replications stress

To map the transcriptional consequences of overexpression of CDC25A, CCNE1 and MYC, RNA sequencing of RPE1-*TP53*^{wt}, RPE1-*TP53*^{mut}, MDA-MB-231, BT549, and HCC-1806, cells was performed both at 48 hours and at 120 hours after oncogene induction to capture gene expression changes provoked by replication stress (Fig. 2A). Subsequently, all cell lines and genetic perturbations were analyzed in a pooled fashion. To this end, the pooled RNAseq dataset was first normalized to remove any possible effects of doxycycline treatment in control cell lines as well as cell line-specific effects (see Supplementary Methods for details). Subsequently, permutational multivariate analysis of variance (PERMANOVA) was performed to identify genes that were differentially expressed upon overexpression of each oncogene (CDC25A, CCNE1, and MYC) across cell lines (Fig. 2B). Induction of CCNE1 or

CDC25A led to significantly differentially upregulated genes ($n=1,330$ and $n=309$, respectively; $p<0.01$), with only 2 genes being downregulated (Fig. 2A, right panel). In contrast, MYC overexpression resulted in a substantial differential gene expression, involving both downregulation ($n=2,576$) and upregulation ($n=935$) of gene expression (Fig. 2A, right panel). Interestingly, expression of 52 genes was found to be commonly upregulated in response to induction of CCNE1, CDC25A, or MYC (Fig. 2A, right panel).

Importantly, gene-set enrichment analysis (GSEA) on metrics obtained from PERMANOVA (see Supplementary Methods for details) revealed that common biological pathways were affected upon induction of CCNE1, CDC25A, or MYC (Fig. 2B), with strong upregulation in expression of MYC targets, and genes involved in cell cycle control and oxidative phosphorylation (Fig. 2B). In contrast, expression of genes related to biological pathways involved in cellular morphology and inflammatory signaling were commonly downregulated (Fig. 2B). Taken together, these results indicate that oncogene overexpression induces distinct yet overlapping gene expression changes, affecting common biological pathways.

Common differential gene expression upon oncogene overexpression between in vitro models and patient samples

To investigate whether the differential gene expression we observed in our cell line models overlaps with patient derived tumor samples with amplification of these oncogenes, we retrieved copy number data and mRNA expression data from The Cancer Genome Atlas (TCGA) (Fig. 3A). TCGA samples were classified into two groups based on the copy number of each of the oncogenes ('amplified' versus 'neutral'). After that, genes that were significantly differentially expressed upon amplification of each oncogene (CDC25A, CCNE1, or MYC) were identified (Fig. 3A). Upon amplification of the three oncogenes, expression of 720 common genes was significantly upregulated (permutation test: $p<1.0\times 10^{-6}$), and that of 597 genes was down-regulated (permutation test: $p<1.0\times 10^{-6}$). GSEA revealed strong upregulation in expression of genes related to MYC targets, cell cycle control, and oxidative phosphorylation (Fig. 3B). In contrast, expression of genes related to immune signatures was commonly downregulated in samples with oncogene amplification (Fig. 3B). The majority of the enrichments in the TCGA data were similar to those obtained by differential expression analysis from the cell line data. Of note, two genesets (i.e., "allograft rejection" and "IL-6/JAK/STAT3 signaling") were only significantly enriched in the analysis with patient derived

tumor samples. This is in line with immune activities in the patient samples not being reflected in cell line models.

Since replication stress can also be caused by oncogenes beyond *MYC*, *CCNE1*, and *CDC25A*, we additionally identified commonly upregulated genes in TCGA tumor samples overexpressing other oncogenes that have been associated with replication stress, including *CCND1*, *MYB*, *MOS*, *KRAS*, *ERBB2*, and *E2F1*^{11,17,25–30}. (Fig. 3C). Analysis of genes that were commonly upregulated upon oncogenes expression in cell line models as well as patient-derived tumor samples revealed six genes (i.e., *NAT10*, *DDX27*, *ZNF48*, *C8ORF33*, *MOCS3*, and *MPP6*) (Fig. 3C, D). Co-functionality analysis of commonly upregulated genes using the GenetICA algorithm (see Supplementary Methods for details) pointed at roles for these genes in ncRNA processing, DNA repair, and ribosome biogenesis (Fig. 3D).

NAT10 protein expression is associated with markers of replication stress in breast cancer samples

To validate the mRNA-based replication stress signature, we selected *NAT10* for immunohistochemical analysis. The acetyltransferase *NAT10* (N-Acetyltransferase-10) has previously been implicated in regulation of various processes, including regulation of the DNA damage response^{31,32} and regulation of translation^{33,34}. *NAT10* is a nuclear protein, with predominant localization to the nucleolus³⁵. Immunohistochemical analysis of *NAT10* staining in a series of breast cancer tissues (n=410), confirmed nuclear localization of *NAT10*, with nucleolar expression in a subset of cancer samples (Fig. 4A). Importantly, a significant variation in protein expression was observed, with triple-negative breast cancer samples showing the highest *NAT10* expression levels (Fig. 4B). Of note, absence or presence of nucleolar localization was not different between breast cancer subgroups (Fig. 4B).

For this breast cancer cohort, we previously reported the presence of p-RPA, a marker of replication stress, and γ H2AX, which reflects DNA breaks, a possible result of stalled replication forks collapse³⁶. In line with *NAT10* expression being part of the oncogene-induced replication stress signature, Spearman correlation analysis showed an association of *NAT10* expression with both p-RPA (R=0.451, p=1.82x10⁻²⁰) and γ H2AX (R=0.304, p=2.95x10⁻⁹) (Table 1A). Subgroup analysis showed that the most significant correlations between *NAT10* expression and pRPA or γ H2AX were observed in ER/PR⁻/HER2⁺ and TNBC samples (Table 1A).

Next, we tested whether NAT10 expression was also correlated to expression of two oncogenes that are frequently amplified in breast cancer, Cyclin E1 and MYC. NAT10 expression was positively correlated to expression of both Cyclin E1 ($R=0.303$, $p=1.86 \times 10^{-9}$) and MYC ($R=0.264$, $p=1.45 \times 10^{-7}$) (Table 1B). Again, associations were most significant in ER/PR-/HER2⁺ and TNBC samples (Table 1B). Combined, these analyses validated NAT10 as part of our oncogene-induced replication stress signature, which is associated with markers of replication stress as well as expression of oncogenes, which are known to induce replication stress.

Landscape of replication stress across tumor types.

To investigate the landscape of oncogene-induced replication stress across cancer types, we used the six-gene signature of oncogene-induced replication stress, as a proxy for oncogene-induced replication stress levels. We applied our signature to RNAseq expression data of 8,862 samples retrieved from TCGA. This dataset represents 27 cancer subtypes as well as non-cancer tissues, and displayed large differences in the oncogene-induced replication stress signature score, with diffuse large B cell lymphoma (DLBCL), ovarian cancer and colorectal carcinoma showing highest scores (Fig. 4A). In line with expectations, normal tissues were among the tissue types with lowest scores (Fig. 4A). These observations are in line with previous reports on these cancer subtypes^{16,37,38}. To confirm the landscape of oncogene-induced replication stress levels in an independent dataset, we retrieved microarray mRNA expression data of 13,912 patient-derived samples from the GEO database (Fig. 4B). A high concordance between the tumor type replication stress levels in TCGA and GEO was observed (Pearson $R=0.77$), indicating that our oncogene-induced replication stress signature captures the level of oncogene-induced replication stress also in a platform-independent fashion.

DISCUSSION

In the present study, we examined transcriptional changes that go along with oncogene-induced replication stress in cell line models and tumor samples to build a gene expression signature of oncogene-induced replication stress. Analysis of a panel of TNBC and non-transformed RPE1 cell lines, combined with analysis of a large dataset of patient-derived cancer samples, yielded a six-gene signature of oncogene-induced replication stress.

Our mRNA-based signature points towards ovarian cancers, colorectal cancers, DLBCLs, TNBCs, cholangiocarcinomas, and esophageal carcinomas having high levels of replication stress. Among these are cancer subtypes that were previously described as prototypical cancers with high levels of replication stress. Specifically, HGSOCs almost invariably have *TP53* mutations (96%), frequently contain high levels of somatic copy number alterations and structural variations, and often have amplification of *MYC* (>30%) and *CCNE1* (>20%)³⁹, both of which have been linked to replication stress and genomic instability^{13,40}. Likewise, TNBCs show biological similarities with HGSOC and also frequently contain somatic *TP53* mutations as well as amplification of *MYC* and *CCNE1*^{41,42}. In good agreement with these data, our immunohistochemical analysis of TNBC samples revealed high levels of p-RPA and γ H2AX, markers of single-stranded DNA and DNA breaks, which have associated with replication stress³⁶.

Cholangiocarcinoma was among the highest-ranked cancer subtypes based on our replication stress signature. These bile duct cancers have not previously been linked to genomic features associated with replication stress. However, recent studies described recurrent alterations in the proto-oncogene *CCND1*, cell cycle regulatory gene *CDKN2A* as well as the chromatin remodelers *ARID1A*, *IDH1/2* and *PBRM1*⁴³, which could underlie DNA replication perturbations^{26,44–47}. Of note, mixed type hepatocellular-cholangiocarcinoma was described to share genetic features with hepatocellular carcinoma⁴⁸, including *CCNE1* amplification, which are causally implicated in hepatocellular carcinogenesis¹⁴.

Also, diffuse large B-cell lymphoma (DLBCL) showed a high score using our gene expression signature. This finding is in line with observations of *MYC* amplification, *ARID1A* mutation and *CDKN2A/B* deletion in DLBCL, with accompanying sensitivity to inhibitors of replication checkpoint kinases^{16,49,50}. Although colorectal cancer (CRC) is not commonly regarded as a tumor type with high levels of oncogene-induced replication stress, CRC scored high in our classifier. This observation is strengthened by earlier observations that CRC subgroups in which pRB is inactivated show activation of the DNA damage response^{11,51}.

Additionally, the chromosomal instability that characterizes CRCs was shown to be linked to replication stress, as judged by slower replication fork progression, and increased numbers of ultrafine anaphase bridges and 53BP1 bodies in G1 cells³⁷.

A gene expression signature may aid in patient selection for drugs that target cancers with high levels of replication stress. Early clinical trials evaluating inhibitors of the ATR, WEE1, and CHK1 checkpoint kinases have shown promising results^{18,52,53}, but not all patients respond and biomarkers defining optimal patient subgroups has been challenging. For instance, *TP53* mutation status was used to select patients for Wee1 inhibitor treatment, but additional features are needed to define patients who will likely benefit^{18,54}. Interestingly, *CCNE1* amplification was among the genetic features that appeared enriched in patients responding to Wee1 inhibitor treatment¹⁸. Focused analysis of oncogene-induced replication stress in these clinical trials is warranted to test whether a more optimal patient selection is achievable.

Materials and methods

Cell lines

hTERT-immortalized human retina epithelial (RPE1), Human embryonic kidney 293 (HEK-293T) and triple-negative breast cancer cell lines MDA-MB-231, BT549, and HCC1806 were obtained from ATCC (#CRL4000, #CRL3216, #HTB26, #HTB122, #CRL2335). RPE1, HEK-293T, and MDA-MB-231 cells were maintained in Dulbecco's Minimum Essential Media (DMEM, Thermofisher), supplemented with 10% (v/v) fetal calf serum (FCS) and 1% penicillin/streptomycin (Gibco). BT-549 and HCC1806 were cultured in Roswell Park Memorial Institute medium (RPMI, Thermofisher), complemented with 10% FCS and 1% penicillin/streptomycin. All cells were grown at 37°C in normoxic conditions (20% O₂ and 5% CO₂). To inactivate *TP53* in RPE1 cells, exon 4 of *TP53* was targeted with CRISPR-CAS9 using the following single guide RNA (sgRNA) (5'-CTGTCATCTTCTGTCCCTTC-3') and *TP53* mutation was confirmed using Sanger sequencing. Additional technical details are provided in *Supplemental Methods*.

DNA fiber analysis

To assess replication dynamics, cell lines were pulse-labeled with CldU and IdU for 20 minutes each. Cells were lysed on top of a microscope slides and DNA fibers were spread by tilting the microscope slide and CldU was detected with rat anti-BrdU (Abcam, ab6326), whereas IdU was detected with mouse anti-BrdU (BD Biosciences, Clone B44). The lengths of 300 IdU tracts were measured per condition using ImageJ software. Statistical analysis was performed using the nonparametric Mann-Whitney U test. Additional technical details are provided in *Supplemental Methods*.

RNA sequencing

Expression of *CDC25A*, *CCNE1*, or *MYC* was induced using 1 µg/mL of doxycycline in biological replicates. Cells were harvested and frozen at -80°C at 48 hours and 120 hours after doxycycline induction. RNA was isolated using the mirVANA kit (Ambion, AM1561). The QuantSeq RNAseq 3' mRNA kit (Lexogen) was employed to generate cDNA libraries for next-generation sequencing (NGS). The libraries were sequenced with 65 base-pair reads on a NextSeq 500 sequencer (Illumina). Additional technical details are provided in *Supplemental Methods*.

Immunohistochemistry

Immunohistochemical analysis was performed on tissue specimens taken before treatment of 558 patients with breast cancer who underwent surgery at the University Medical Center Groningen (UMCG) as described previously³⁶. Immunohistochemistry was performed with primary antibody against NAT10 (1:500; rabbit, #ab194297, clone EPR18663; Abcam, Cambridge, UK) for 1 hour at room temperature. NAT10 expression was categorized according to percentages of cells that showed staining and on the intensity of staining. Staining intensity was scored in three categories: 0 (negative), 1 (medium), and 2 (high). To calculate the score for each core, the percentage of cells in each group was multiplied by their intensity score, resulting in a range from 0 to 200 points. In addition to NAT10 expression scores, the tumors were scored for the presence of nucleolar NAT10 localization. Next, the scores from each case and staining were averaged and considered for analysis. The status of ER, PR, and HER2 was determined according to the guidelines of the American Society of Clinical Oncology/College of American Pathologists by counting at least 100 cells. Of 558 cases, 410 cases contained complete clinical data and evaluable immunohistochemical staining for NAT10 expression and were included for statistical analysis. Analyses were performed on the 410 cases as well as on patient subgroups based on hormone receptor status and HER2 expression. Additional technical details are provided in *Supplemental Methods*.

Data availability

All sequencing data have been deposited at the GEO archive of NCBI.

Funding

This work was financially supported by grant from the Netherlands Organization for Scientific Research (NWO-VIDI #917.13334 to M.A.T.M.v.V. and NWO-VENI #916-16025 to R.S.N.F.), the Dutch Cancer Society (RUG 2013-5960 to R.S.N.F) and from the European Research Council (ERC-Consolidator grant “TENSION” to M.A.T.M.v.V.).

Acknowledgments

We thank members of the Medical Oncology laboratory for fruitful discussions.

Conflict of Interest

M.A.T.M.v.V. has acted on the Scientific Advisory Board of Repare Therapeutics, which is unrelated to this work. The other authors declare no conflict of interest.

References

1. Koboldt DC, Fulton RS, McLellan MD, et al. Comprehensive molecular portraits of human breast tumours. *Nature*. Published online 2012. doi:10.1038/nature11412
2. Curtis C, Shah P, Chin S-F, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. Published online 2012.
3. Gerlinger M, Swanton C. How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. *Br J Cancer*. Published online 2010. doi:10.1038/sj.bjc.6605912
4. Bell D, Berchuck A, Birrer M, et al. Integrated genomic analyses of ovarian carcinoma. *Nature*. Published online 2011. doi:10.1038/nature10166
5. Bergamaschi A, Kim YH, Wang P, et al. Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. *Genes Chromosomes Cancer*. Published online 2006. doi:10.1002/gcc.20366
6. Schoonen PM, Guerrero I, Llobet S, van Vugt MATM. Replication stress: Driver and therapeutic target in genomically unstable cancers. In: *Advances in Protein Chemistry and Structural Biology*. ; 2019. doi:10.1016/bs.apcsb.2018.10.006
7. Zeman MK, Cimprich KA. Causes and consequences of replication stress. *Nat Cell Biol*. Published online 2014. doi:10.1038/ncb2897
8. Bell SP, Dutta A. DNA replication in eukaryotic cells. *Annu Rev Biochem*. Published online 2002. doi:10.1146/annurev.biochem.71.110601.135425
9. Bester AC, Roniger M, Oren YS, et al. Nucleotide deficiency promotes genomic instability in early stages of cancer development. *Cell*. Published online 2011. doi:10.1016/j.cell.2011.03.044
10. Halazonetis TD, Gorgoulis VG, Bartek J. An oncogene-induced DNA damage model for cancer development. *Science (80-)*. Published online 2008. doi:10.1126/science.1140735
11. Bartkova J, Hořejší Z, Koed K, et al. DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature*. Published online 2005. doi:10.1038/nature03482

12. Macheret M, Halazonetis TD. Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. *Nature*. Published online 2018. doi:10.1038/nature25507
13. Neelsen KJ, Zanini IMY, Herrador R, Lopes M. Oncogenes induce genotoxic stress by mitotic processing of unusual replication intermediates. *J Cell Biol*. Published online 2013. doi:10.1083/jcb.201212058
14. Aziz K, Limzerwala JF, Sturmlechner I, et al. Ccne1 Overexpression Causes Chromosome Instability in Liver Cells and Liver Tumor Development in Mice. *Gastroenterology*. Published online 2019. doi:10.1053/j.gastro.2019.03.016
15. López-Contreras AJ, Gutierrez-Martinez P, Specks J, Rodrigo-Perez S, Fernandez-Capetillo O. An extra allele of Chk1 limits oncogene-induced replicative stress and promotes transformation. *J Exp Med*. Published online 2012. doi:10.1084/jem.20112147
16. Young LA, O'Connor LO, de Renty C, et al. Differential activity of ATR and Wee1 inhibitors in a highly sensitive subpopulation of DLBCL linked to replication stress. *Cancer Res*. Published online 2019. doi:10.1158/0008-5472.CAN-18-2480
17. Bartkova J, Rezaei N, Liontos M, et al. Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature*. Published online 2006. doi:10.1038/nature05268
18. Leijen S, Van Geel RMJM, Sonke GS, et al. Phase II study of WEE1 inhibitor AZD1775 plus carboplatin in patients with tp53-mutated ovarian cancer refractory or resistant to first-line therapy within 3 months. *J Clin Oncol*. Published online 2016. doi:10.1200/JCO.2016.67.5942
19. Kok YP, Llobet SG, Schoonen PM, et al. Overexpression of Cyclin E1 or Cdc25A leads to replication stress, mitotic aberrancies, and increased sensitivity to replication checkpoint inhibitors. *Oncog 2020 910*. Published online 2020.
20. Heijink AM, Blomen VA, Bisteau X, et al. A haploid genetic screen identifies the G1/S regulatory machinery as a determinant of Wee1 inhibitor sensitivity. *Proc Natl Acad Sci U S A*. Published online 2015. doi:10.1073/pnas.1505283112
21. Aarts M, Sharpe R, Garcia-Murillas I, et al. Forced mitotic entry of S-phase cells as a therapeutic strategy induced by inhibition of WEE1. *Cancer Discov*. Published online 2012. doi:10.1158/2159-8290.CD-11-0320
22. Wengner AM, Siemeister G, Lucking U, et al. The novel ATR inhibitor BAY 1895344 is efficacious as monotherapy and combined with DNA damage-inducing or repair-

- compromising therapies in preclinical cancer models. *Mol Cancer Ther.* Published online 2020. doi:10.1158/1535-7163.MCT-19-0019
23. Sakurikar N, Thompson R, Montano R, Eastman A. A subset of cancer cell lines is acutely sensitive to the Chk1 inhibitor MK-8776 as monotherapy due to CDK2 activation in S phase. *Oncotarget.* Published online 2016. doi:10.18632/oncotarget.6364
 24. Ferrao PT, Bukczynska EP, Johnstone RW, McArthur GA. Efficacy of CHK inhibitors as single agents in MYC-driven lymphoma cells. *Oncogene.* Published online 2012. doi:10.1038/onc.2011.358
 25. Shimura T, Ochiai Y, Noma N, Oikawa T, Sano Y, Fukumoto M. Cyclin D1 overexpression perturbs DNA replication and induces replication-associated DNA double-strand breaks in acquired radioresistant cells. *Cell Cycle.* Published online 2013. doi:10.4161/cc.23719
 26. Aggarwal P, Lessie MD, Lin DI, et al. Nuclear accumulation of cyclin D1 during S phase inhibits Cul4-dependent Cdt1 proteolysis and triggers p53-dependent DNA rereplication. *Genes Dev.* Published online 2007. doi:10.1101/gad.1586007
 27. Maya-Mendoza A, Ostrakova J, Kosar M, et al. Myc and Ras oncogenes engage different energy metabolism programs and evoke distinct patterns of oxidative and DNA replication stress. *Mol Oncol.* Published online 2015. doi:10.1016/j.molonc.2014.11.001
 28. Di Micco R, Fumagalli M, Cicalese A, et al. Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature.* Published online 2006. doi:10.1038/nature05327
 29. Carlos AR, Escandell JM, Kotsantis P, et al. ARF triggers senescence in Brca2-deficient cells by altering the spectrum of p53 transcriptional targets. *Nat Commun.* Published online 2013. doi:10.1038/ncomms3697
 30. Kanu N, Cerone MA, Goh G, et al. DNA replication stress mediates APOBEC3 family mutagenesis in breast cancer. *Genome Biol.* Published online 2016. doi:10.1186/s13059-016-1042-9
 31. Liu HY, Liu YY, Yang F, et al. Acetylation of MORC2 by NAT10 regulates cell-cycle checkpoint control and resistance to DNA-damaging chemotherapy and radiotherapy in breast cancer. *Nucleic Acids Res.* Published online 2020. doi:10.1093/nar/gkaa130
 32. Liu X, Tan Y, Zhang C, et al. NAT 10 regulates p53 activation through acetylating p53 at K120 and ubiquitinating Mdm2. *EMBO Rep.* Published online 2016.

- doi:10.15252/embr.201540505
33. Arango D, Sturgill D, Alhusaini N, et al. Acetylation of Cytidine in mRNA Promotes Translation Efficiency. *Cell*. Published online 2018. doi:10.1016/j.cell.2018.10.030
 34. Dominissini D, Rechavi G. N4-acetylation of Cytidine in mRNA by NAT10 Regulates Stability and Translation. *Cell*. Published online 2018. doi:10.1016/j.cell.2018.11.037
 35. Shen Q, Zheng X, McNutt MA, et al. NAT10, a nucleolar protein, localizes to the midbody and regulates cytokinesis and acetylation of microtubules. *Exp Cell Res*. Published online 2009. doi:10.1016/j.yexcr.2009.03.007
 36. Guerrero Llobet S, van der Vegt B, Jongeneel E, et al. Cyclin E expression is associated with high levels of replication stress in triple-negative breast cancer. *npj Breast Cancer*. Published online 2020. doi:10.1038/s41523-020-00181-w
 37. Burrell RA, McClelland SE, Endesfelder D, et al. Replication stress links structural and numerical cancer chromosomal instability. *Nature*. Published online 2013. doi:10.1038/nature11935
 38. Pillay N, Tighe A, Nelson L, et al. DNA Replication Vulnerabilities Render Ovarian Cancer Cells Sensitive to Poly(ADP-Ribose) Glycohydrolase Inhibitors. *Cancer Cell*. Published online 2019. doi:10.1016/j.ccell.2019.02.004
 39. Campbell PJ, Getz G, Korbel JO, et al. Pan-cancer analysis of whole genomes. *Nature*. Published online 2020. doi:10.1038/s41586-020-1969-6
 40. Murga M, Campaner S, Lopez-Contreras AJ, et al. Exploiting oncogene-induced replicative stress for the selective killing of Myc-driven tumors. *Nat Struct Mol Biol*. Published online 2011. doi:10.1038/nsmb.2189
 41. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. Published online 2012. doi:10.1038/nature10983
 42. Banerji S, Cibulskis K, Rangel-Escareno C, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature*. Published online 2012. doi:10.1038/nature11154
 43. Farshidfar F, Zheng S, Gingras MC, et al. Integrative Genomic Analysis of Cholangiocarcinoma Identifies Distinct IDH-Mutant Molecular Profiles. *Cell Rep*. Published online 2017. doi:10.1016/j.celrep.2017.02.033
 44. Gadhikar MA, Zhang J, Shen L, et al. CDKN2A/p16 deletion in head and neck cancer cells is associated with cdk2 activation, replication stress, and vulnerability to CHK1 inhibition. *Cancer Res*. Published online 2018. doi:10.1158/0008-5472.CAN-17-2802

45. Williamson CT, Miller R, Pemberton HN, et al. ATR inhibitors as a synthetic lethal therapy for tumours deficient in ARID1A. *Nat Commun*. Published online 2016. doi:10.1038/ncomms13837
46. Inoue S, Li WY, Tseng A, et al. Mutant IDH1 Downregulates ATM and Alters DNA Repair and Sensitivity to DNA Damage Independent of TET2. *Cancer Cell*. Published online 2016. doi:10.1016/j.ccell.2016.05.018
47. Espana-Agusti J, Warren A, Chew SK, Adams DJ, Matakidou A. Loss of PBRM1 rescues VHL dependent replication stress to promote renal carcinogenesis. *Nat Commun*. Published online 2017. doi:10.1038/s41467-017-02245-1
48. Joseph NM, Tsokos CG, Umetsu SE, et al. Genomic profiling of combined hepatocellular-cholangiocarcinoma reveals similar genetics to hepatocellular carcinoma. *J Pathol*. Published online 2019. doi:10.1002/path.5243
49. De Jong MRW, Visser L, Huls G, et al. Identification of relevant drugable targets in diffuse large B-cell lymphoma using a genome-wide unbiased CD20 guilt-by-association approach. *PLoS One*. Published online 2018. doi:10.1371/journal.pone.0193098
50. Reddy A, Zhang J, Davis NS, et al. Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell*. Published online 2017. doi:10.1016/j.cell.2017.09.027
51. Tort F, Bartkova J, Sehested M, Ørntoft T, Lukas J, Bartek J. Retinoblastoma pathway defects show differential ability to activate the constitutive DNA damage response in human tumorigenesis. *Cancer Res*. Published online 2006. doi:10.1158/0008-5472.CAN-06-2178
52. Konstantinopoulos PA, Cheng SC, Wahner Hendrickson AE, et al. Berzosertib plus gemcitabine versus gemcitabine alone in platinum-resistant high-grade serous ovarian cancer: a multicentre, open-label, randomised, phase 2 trial. *Lancet Oncol*. Published online 2020. doi:10.1016/S1470-2045(20)30180-7
53. Italiano A, Infante JR, Shapiro GI, et al. Phase I study of the checkpoint kinase 1 inhibitor GDC-0575 in combination with gemcitabine in patients with refractory solid tumors. *Ann Oncol*. Published online 2018. doi:10.1093/annonc/mdy076
54. Rajeshkumar N V., De Oliveira E, Ottenhof N, et al. MK-1775, a potent Wee1 inhibitor, synergizes with gemcitabine to achieve tumor regressions, selectively in p53-deficient pancreatic cancer xenografts. *Clin Cancer Res*. Published online 2011. doi:10.1158/1078-0432.CCR-10-2580

Table 1A Spearman correlation analysis of γ -H2AX and p-RPA versus NAT10 expression

Variable	BC Subgroup				
	Combined cohort	ER/PR ⁺ HER2 ⁻	ER/PR ⁺ HER2 ⁺	ER/PR ⁻ HER2 ⁺	TNBC
	n=502	n=211	n=120	n=28	n=143
	NAT10				
γ -H2AX	n=330	n=132	n=78	n=20	n=100
Correlation	0.316	0.275	0.015	0.614	0.297
P-value	4.23x10 ⁻⁹	0.001	0.898	0.004	0.003
p-RPA	n=330	n=132	n=78	n=20	n=100
Correlation	0.454	0.280	0.256	0.498	0.483
P-value	3.42x10 ⁻¹⁸	0.001	0.023	0.026	2.00x10 ⁻¹⁰

Table 1B. Spearman correlation analysis of Cyclin E, MYC versus NAT10 expression

Variable	BC Subgroup				
	Combined cohort	ER/PR ⁺ HER2 ⁻	ER/PR ⁺ HER2 ⁺	ER/PR ⁻ HER2 ⁺	TNBC
	n=502	n=211	n=120	n=28	n=143
	NAT10				
Cyclin E	n=330	n=132	n=78	n=20	n=100
Correlation	0.324	0.136	0.218	0.625	0.299
P-value	1.76x10 ⁻⁹	0.120	0.141	0.003	0.002
MYC	n=330	n=132	n=78	n=20	n=100
Correlation	0.282	0.171	0.170	0.238	0.272
P-value	1.91x10 ⁻⁷	0.050	0.137	0.311	0.006

Table 1: Spearman rank correlation analysis of γ H2AX, p-RPA, and oncogene expression versus NAT10 expression in the study population.

(A) Association analysis between γ H2AX, p-RPA and NAT10 expression in the combined cohort (n=330) and in breast cancer subgroups. (B) Association analysis between oncogene expression (Cyclin E and MYC) and NAT10 expression in the combined cohort (n=330) and in breast cancer subgroups.

Figure legends

Figure 1: Overexpression of *CDC25A*, *CCNE1* or *MYC* leads to replication stress.

(A) Indicated RPE1-*TP53*^{wt} cell lines were treated with doxycycline for 48 hours. Immunoblotting was performed for *CDC25A*, *CCNE1*, p53, and β -Actin. (B) Cells were treated with doxycycline as described for panel A and subsequently pulse-labeled for 20 minutes with CldU (25 μ M) and 20 minutes with IdU (250 μ M). Representative DNA fibers from RPE1-*TP53*^{wt} cells are shown. Scale bar represents 10 μ m. (C-G) Quantification of IdU DNA fiber lengths, as described in panel B. Per condition, 300 fibers from RPE1-*TP53*^{wt} cells were analyzed. P values were calculated using the Mann-Whitney U test for RPE1-*TP53*^{wt} (panel C), RPE1-*TP53*^{mut} (panel D), MDA-MB-231 (panel E), BT549 (panel F) and HCC1806 (panel G).

Figure 2: Overexpression of *CDC25A*, *CCNE1*, or *MYC* leads to upregulation of 52 genes.

(A) RNAseq data of 5 cell lines were corrected for cell line-specific and doxycycline treatment effects on gene expression. Subsequently, PERMANOVA was used to identify common differentially expressed genes upon oncogene expression. 52 genes were found to be commonly upregulated whereas no genes were found to be commonly downregulated in response to induction of *CCNE1*, *CDC25A* or *MYC* (B) Gene Set Enrichment Analysis using two-sided Welch's t-test for the MSigDB Hallmark collection. An orange box indicates enrichment for upregulated genes due to overexpression of oncogenes in corresponding cell lines, and a green box indicates enrichment for downregulated genes.

Figure 3: Common differential gene expression of 6 genes upon oncogene overexpression between *in vitro* models and patient samples.

(A) TCGA RNAseq samples were queried for amplification of *CCNE1*, *CDC25A* or *MYC*. Expression of 720 genes was found to be commonly upregulated, whereas 597 genes were found to be commonly downregulated in response to amplification of *CCNE1*, *CDC25A* or *MYC*. (B) Gene Set Enrichment Analysis using two-sided Welch's t-test for the MSigDB Hallmark collection. An orange box indicates enrichment for upregulated genes due to overexpression of oncogenes in corresponding cell lines, whereas a green box indicates enrichment for downregulated genes. (C) TCGA RNAseq samples were queried for amplification of *CCND1*, *MYB*, *MOS*, *KRAS*, *ERBB2* and *E2F1*. The resulting commonly

upregulated genes were overlaid with upregulated genes identified upon overexpression of *CCNE1*, *CDC25A* or *MYC* in cell lines and upregulated genes in TCGA samples with *CCNE1*, *CDC25A* or *MYC* amplifications, resulting in 6 commonly upregulated genes **(D)** Co-functionality analysis of commonly upregulated genes using the GenetICA algorithm.

Figure 4: Immunohistochemical analysis of NAT10 in breast cancer patients.

(A) Representative staining of NAT10 in breast cancer patient samples (n=410). Scale bar represents 100 μ m. **(B)** Patients from the combined cohort (n=410) and breast cancer subgroups ER/PR⁺HER2⁻ (n=164), ER/PR⁺HER2⁺ (n=95), ER/PR⁻HER2⁺ (n=21) and TNBC (n=130) were analyzed. Tumor tissue was immunohistochemically scored for expression of NAT10 and NAT10 nucleoli **(C)**, indicated P values were calculated using Mann-Whitney U test. Interquartile ranges are displayed, dashed red line represents the median score from the combined cohort and outliers are shown as dots.

Figure 5: The landscape of oncogene-induced replication stress across cancer types.

(A) Distribution of the replication stress signature across 27 cancer types as well as non-cancerous tissues of patient samples from TCGA. **(B)** Scatter plot of replication stress signature for 27 cancer types as well as non-cancerous tissues of patient samples from TCGA (y axis) and GEO (x axis).

Figure 1

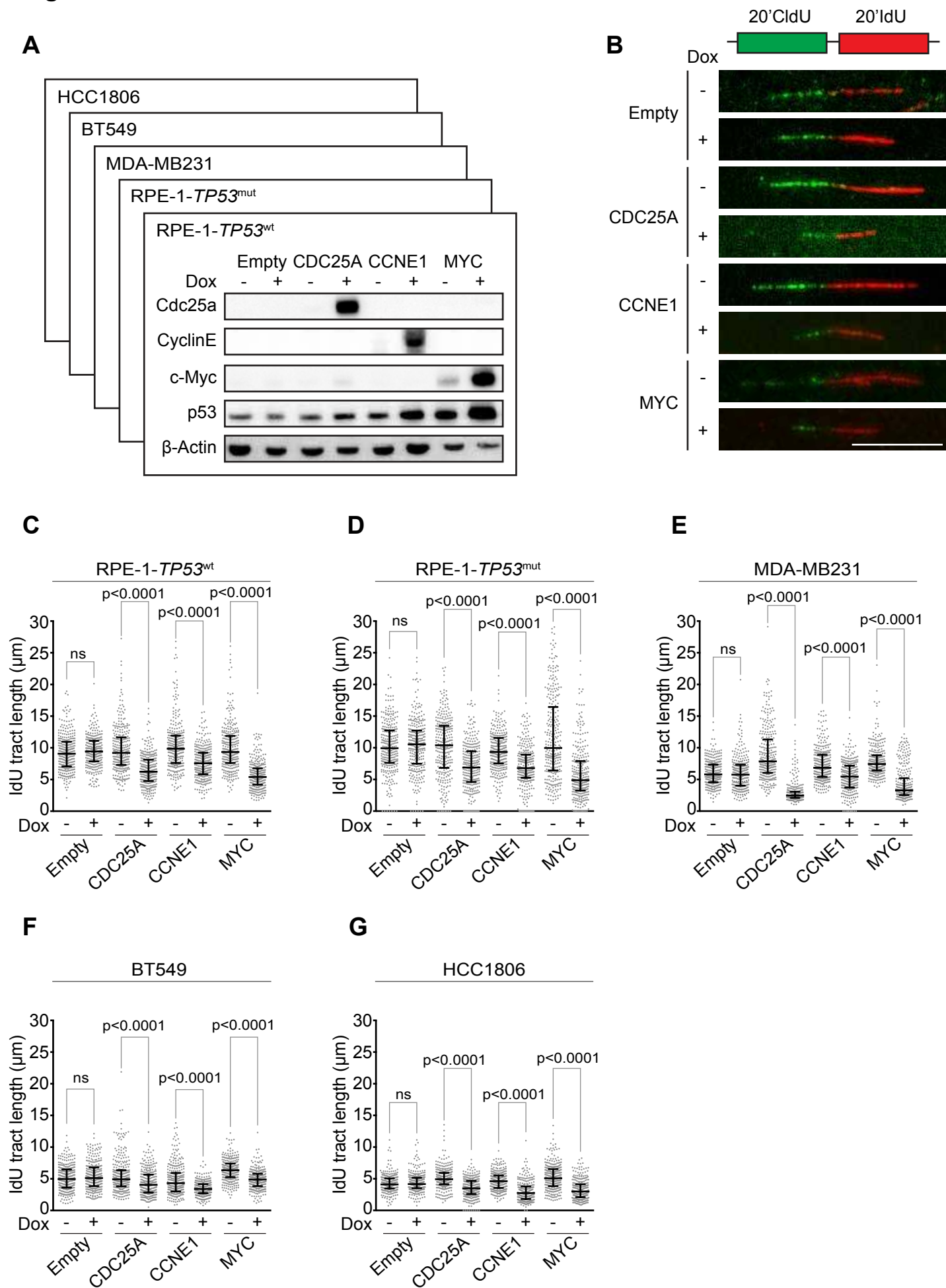


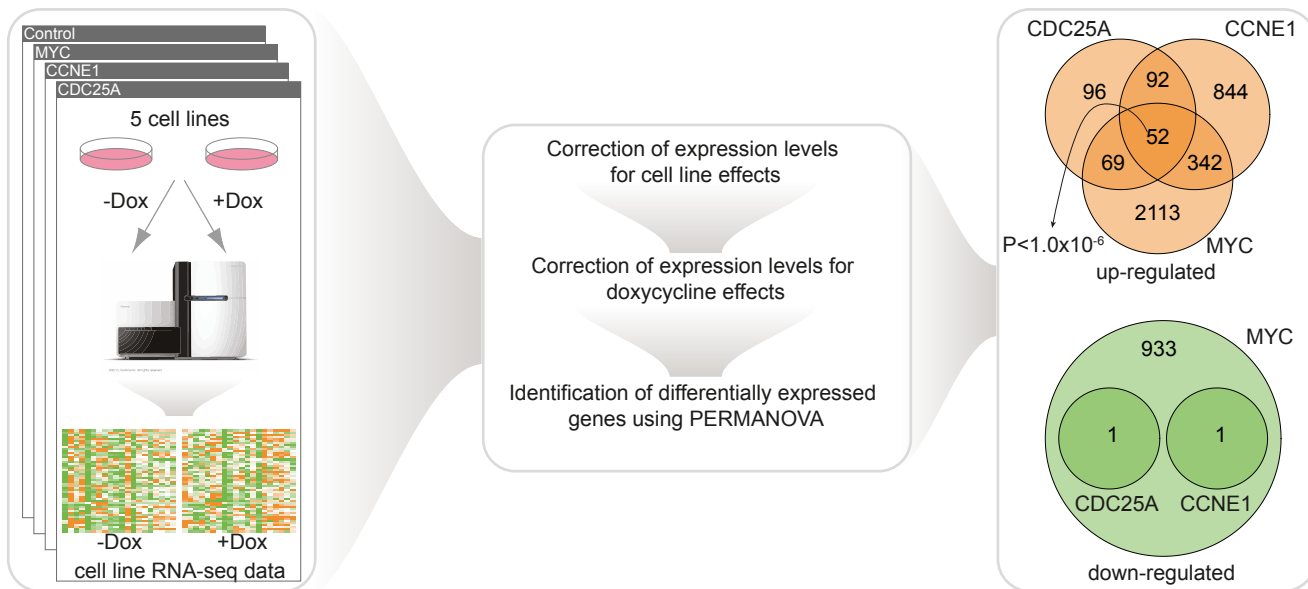
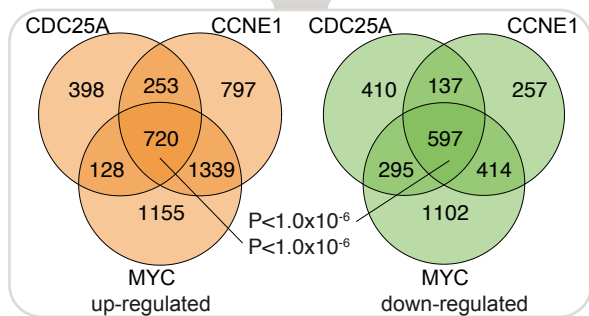
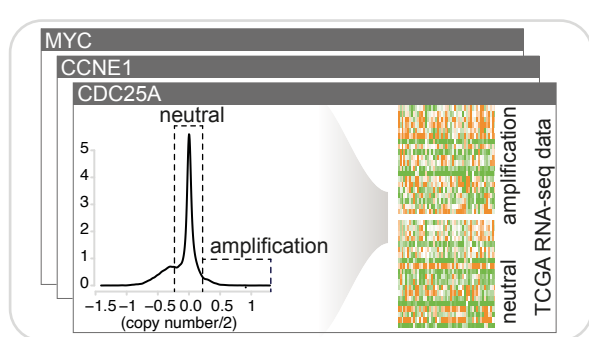
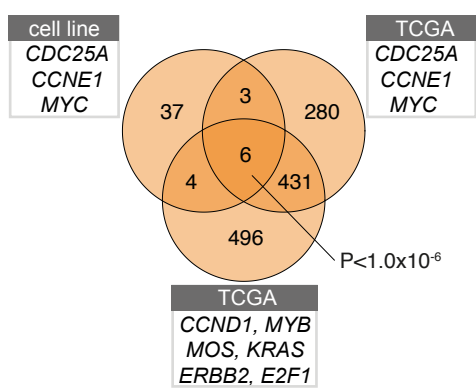
Figure 2**A****B**

Figure 3

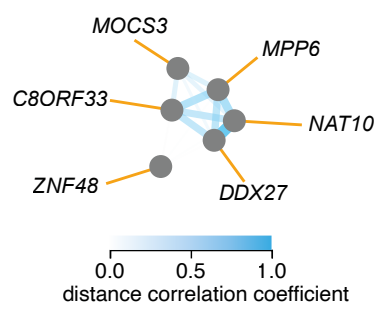
A



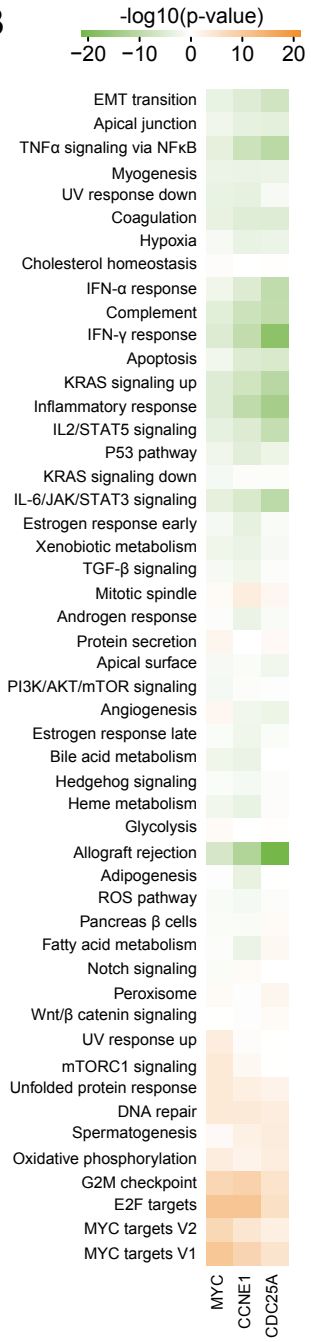
C



D



B



- Z score Enriched gene set of selected genes
- 6.97 NCRNA PROCESSING
 - 6.93 RIBONUCLEOPROTEIN COMPLEX B
 - 6.89 RIBOSOME BIOGENESIS
 - 6.32 DNA REPAIR
 - 6.32 AMIDE BIOSYNTHETIC PROCESS
 - 6.24 ORGANELLE FISSION
 - 5.96 MRNA PROCESSING
 - 5.96 RRNA METABOLIC PROCESS
 - 5.64 RNA SPLICING
 - 5.51 DNA REPLICATION

Figure 4

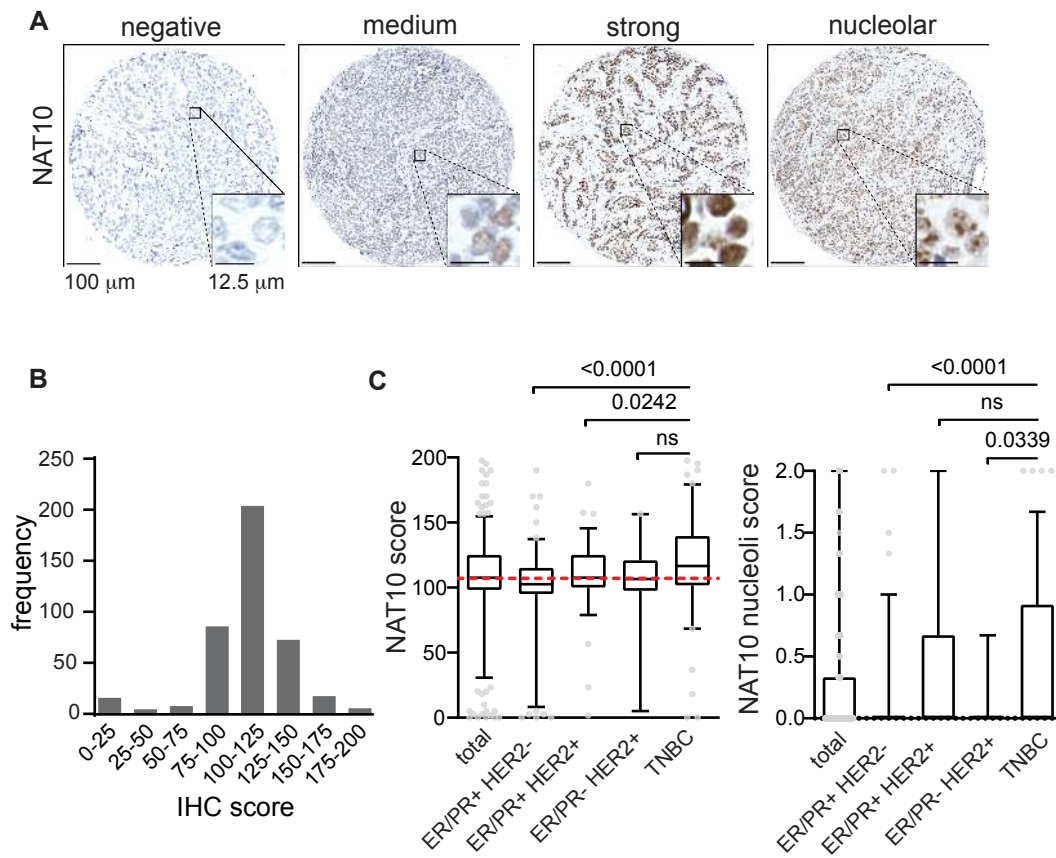


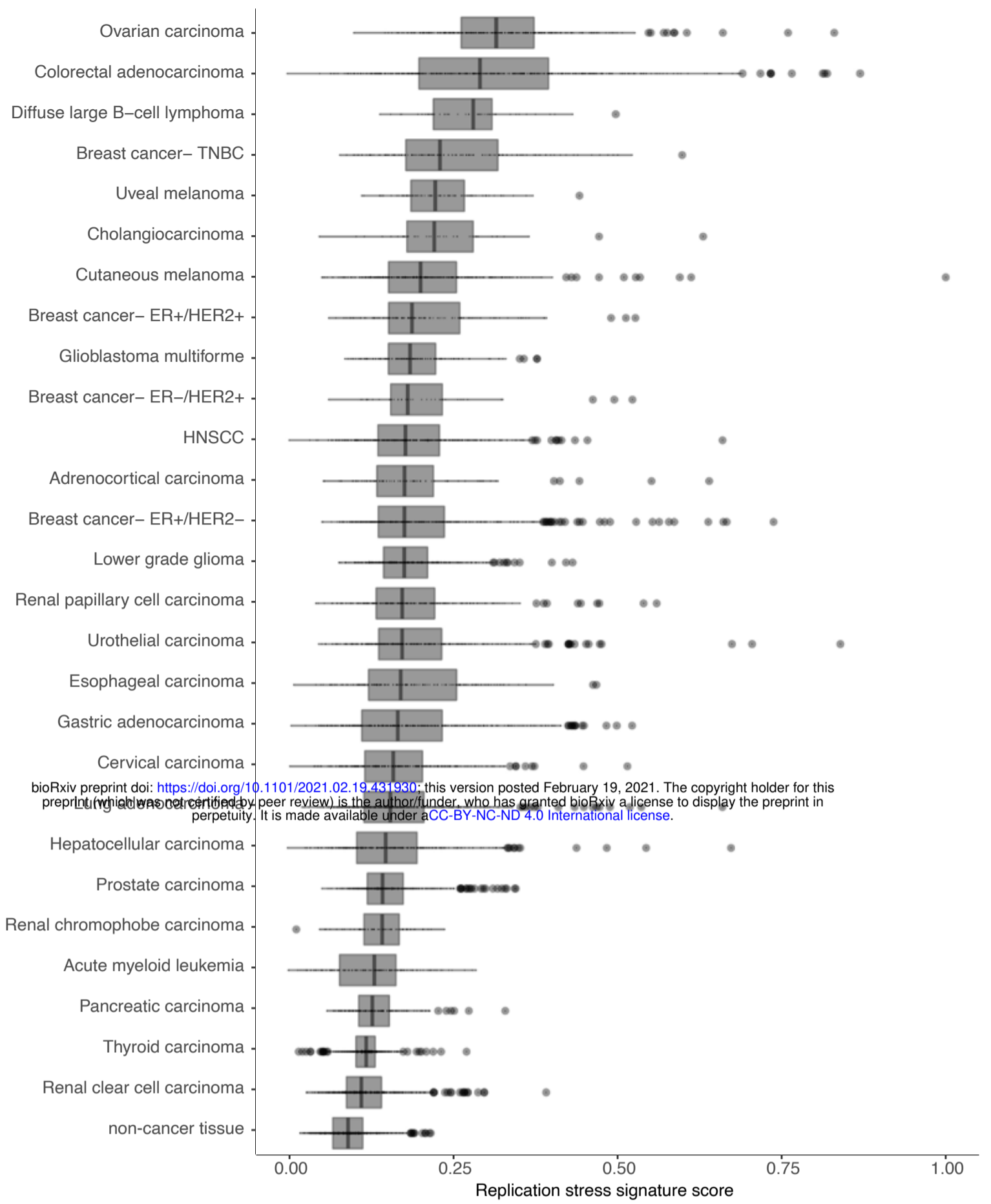
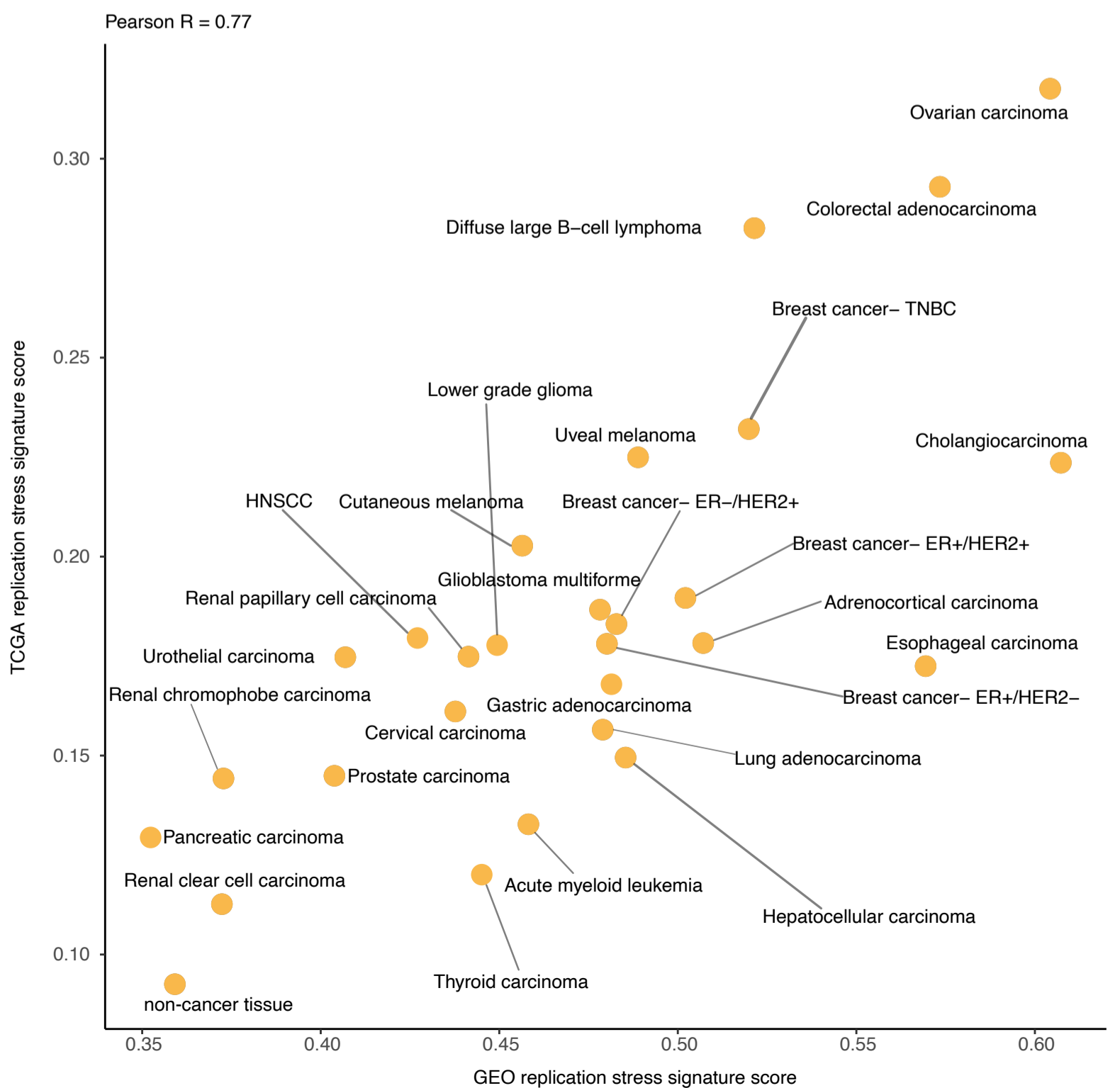
Figure 5**A****B**

Table 1A Spearman correlation analysis of γ -H2AX and p-RPA versus NAT10 expression

Variable	BC Subgroup				
	Combined cohort	ER/PR ⁺ HER2 ⁻	ER/PR ⁺ HER2 ⁺	ER/PR ⁻ HER2 ⁺	TNBC
	n=502	n=211	n=120	n=28	n=143
	NAT10				
γ -H2AX	n=330	n=132	n=78	n=20	n=100
Correlation	0.316	0.275	0.015	0.614	0.297
P-value	4.23*10 ⁻⁹	0.001	0.898	0.004	0.003
p-RPA	n=330	n=132	n=78	n=20	n=100
Correlation	0.454	0.280	0.256	0.498	0.483
P-value	3.42*10 ⁻¹⁸	0.001	0.023	0.026	2.00*10 ⁻¹⁰

Table 1B. Spearman correlation analysis of Cyclin E, MYC versus NAT10 expression

Variable	BC Subgroup				
	Combined cohort	ER/PR ⁺ HER2 ⁻	ER/PR ⁺ HER2 ⁺	ER/PR ⁻ HER2 ⁺	TNBC
	n=502	n=211	n=120	n=28	n=143
	NAT10				
Cyclin E	n=330	n=132	n=78	n=20	n=100
Correlation	0.324	0.136	0.218	0.625	0.299
P-value	1.76*10 ⁻⁹	0.120	0.141	0.003	0.002
MYC	n=330	n=132	n=78	n=20	n=100
Correlation	0.282	0.171	0.170	0.238	0.272
P-value	1.91*10 ⁻⁷	0.050	0.137	0.311	0.006