R-loop editing by DNA cytosine deaminase APOBEC3B determines the activity of estrogen receptor enhancers

Chi Zhang[1,4], Yu-jing Lu[3#], Bingjie Chen[1,5#], Zhiyan Bai[4], Alexia Hervieu[1], Marco P. Licciardello[1], Mei Wang[2], Costas Mitsopoulos[1], Bissan Al-Lazikani[1], Marcello Totorici[1], Olivia W. Rossanese[1*], Paul Workman[1*], Paul A. Clarke[1*]

1. Centre for Cancer Drug Discovery, the Institute of Cancer Research, London, United Kingdom
2. Shanghai Institute of Biological Products, Shanghai, China
3. Institute of Biomedical and Pharmaceutical Sciences, Guangdong University of Technology, Guangzhou, China
4. Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China
5. GMU-GIBH Joint School of Life Sciences, Guangzhou Medical University, Guangzhou, Guangdong, China.

Running Title: APOBEC3B regulates estrogen receptor activity by R-loop editing

#: Equal secondary contributions

*:Correspondence: olivia.rossanese@icr.ac.uk, paul.workman@icr.ac.uk and paul.clarke@icr.ac.uk

Summary

Estrogen receptor (ER) activation results in the formation of DNA double strand breaks (DSB), which promote genomic instability and tumour heterogeneity in ER-positive breast cancers. The single-stranded DNA (ssDNA) cytosine deaminase APOBEC3B (A3B) regulates ER activity by inducing DSB at ER enhancers. To delineate how A3B recognises its substrates and unveil the underlying mechanism leading to the formation of ER-induced DSB, we sampled A3B-mediated deamination sites using whole genome sequencing in a human breast cancer cell model lacking base excision repair function. Our genome-wide analysis revealed that C>U conversions carried out by A3B in R-loop structures are processed into DSB in the vicinity of ER promoters or enhancers. A mechanism which required both the processing of A3B-editing sites and R-loops by distinct DNA damage repair mechanisms. In addition, using BioID-enabled mass-spectroscopy proteomics, we identified TDRD3 as a key A3B-binding partner directing the activity of A3B to ER-induced R-loops. This study suggests a function for

A3B in sustaining tumour evolution as an adaptive response at the transcriptional and epigenetic level and supports A3B as a promising target to control ER activity in cancer.

Keywords:

APOBEC3B; estrogen receptor; R-loop; TDRD3; DNA deamination; double-strand break; DNA damage repair

Introduction

Estrogen (17-β E2; E2) is a hormone that is essential for the development and maintenance of the human reproductive system and secondary sexual characteristics. Prolonged exposure to E2 contributes to the development of breast, ovarian and uterine cancers[1]. E2 binds and activates the two estrogen receptor (ER) isoforms, ERα (*ESR1*) and β (*ESR2*) which are expressed in different tissues, but can form both homo- and heterodimers when co-expressed[2]. Upon E2 binding, ER translocates to the nucleus and acts as a key transcription factor by recognising E2 responsive elements (EREs) in genomic DNA. Translocation of ER to the nucleus and binding of ER to EREs leads to altered gene expression of thousands of downstream genes[3]. The gene set regulated by ER activation encompasses genes encoding essential reproductive functions, including cell proliferation and survival, and is linked to cancer development when ER is inappropriately activated[4]. In addition to this role as a regulator of gene expression, ER can also promote the development of cancer by inducing DNA damage[5]. In ER-expressing breast epithelial cells, short-term stimulation by E2 triggers the formation of DNA double-strand breaks (DSBs)[6, 7], which contributes to chromosome instability and aneuploidy[8]. Increased genomic damage is particularly apparent in ER-positive breast cancers, where there is a high degree of chromosome instability characterized by mutations, gene copy number alterations and recombinations[9, 10]. It has been hypothesised that ER-induced DNA damage is pivotal to the activity of ER since the DNA repair pathway elements not only participate and promote transcription activation, as exemplified by transcription-coupled nucleotide excision repair (TC-NER) factors[11, 12], but also foster favourable conditions for transcription[13] such as DNA demethylation[14], chromatin remodelling [15] and enhancer-associated RNA (eRNA) synthesis[16].

To date, several models have been proposed to account for the generation of E2-induced DSBs, including: the action of DNA topoisomerase IIβ[17], formation and processing of co-transcriptional R-loops[7], and DNA damage triggered by deamination editing[6]. In recent years there has been particular interest in the role of apolipoprotein B mRNA-editing enzyme

catalytic polypeptide-like 3 (APOBEC3B, A3B) and its role in multiple human cancers[18]. A3B is a member of the APOBEC3 enzyme family, which comprises seven closely related DNA deaminases that catalyse cytosine-to-uracil (C>U) editing of single-stranded DNA (ssDNA)[19] or RNA. In addition to their normal functions as immune defences against DNA viruses and transposons [20, 21], APOBEC3 catalytic activity has been reported to drive genetic heterogeneity in breast cancer by inducing C>T transitions and C>G transversions at 5'TCW motifs (W = A or T)[9, 22, 23]. Although mounting evidence is pointing APOBEC3A, a highly homologous enzyme to C-terminal A3B, as the main contributor to APOBEC-driven mutation burden and cancer evolution[24], A3B-driven cytidine deamination still has a prominent role in cancer due to two facts. First, as the only nuclear-localised member of the APOBEC3 family, elevated A3B expression is reported in >50% of primary breast cancers[22, 23]; and secondly, A3B binds directly to ERα and edits ER binding sites, which triggers DNA damage response (DDR) and are then processed by DNA base excision repair (BER) and non-homologous end joining (NHEJ) into DSBs[6]. The right amount of DSBs, together with the subsequent epigenetic modifications, determines the transactivation activity of ER[6]. Loss of functional A3B ablates ER-driven cell proliferation and prolongs the therapeutic benefit of tamoxifen in ER-positive xenograft models[25]. These data support A3B as a regulator of ER, and also as a promising target to tackle drug resistance in ER-targeting therapies.

Elevated transcriptional activity leads to the formation of chromatin-associated R-loop structures, which are formed by re-annealing of nascent RNA with the transcribed strand of DNA and results in a three-stranded structure consisting of a DNA:RNA hybrid and a displaced ssDNA strand[26]. Previous work has shown that clearance of co-transcriptionally formed R-loops by DDR machinery is a main source of DSBs caused by ER activation[7]. Since the displaced ssDNA in R-loops could serve as a substrate for A3B, it has been proposed that A3B could be involved in R-loop-associated functions. This hypothesis is supported by the use of APOBEC-Cas9 engineered system as tools to edit R-loop DNA cytosines[27], and the fact that activation-induced cytidine deaminase (AID), a member of the APOBEC/AID

deaminases family, edits R-loops to facilitate antibody gene hypermutation and class switch recombination in B lymphocytes[28]. More recently, A3B has been found to influence R-loop homeostasis[29]. However, despite these endeavours, it is unknown whether R-loop editing by A3B is directly associated with its function as an ER coactivator.

An ideal way to address the potential role of R-loop editing by A3B on the co-activation of ER-regulated transcription is to sample A3B editing sites on a large scale using ER-dependent models. This approach would reveal potential hot spots of A3B editing, which could be used to infer the mechanism that governs A3B-dependent DNA editing and point to how A3B regulates ER activity. Large-scale whole-genome sequencing (WGS) studies performed on cell lines and primary samples have identified sites of APOBEC editing[10, 30]. However, these studies may be confounded by a 'survivorship bias' where the detected A3B-edited sites are limited to those that were not repaired by DDR, as it is known that BER removes U:G mispairing resulting from A3B editing[24, 31]. This caveat may also affect studies employing A3B-overexpression cell models[32-35]. To circumvent this, we sampled A3B editing sites using an engineered BER-deficient breast cancer cellular model and used a multi-omic approach to elucidate the role of A3B in the molecular mechanism of ER-regulated gene expression.

Results

Detection of A3B deamination sites using a BER-deficient human breast cancer cell model

In order to capture A3B deamination sites that may have a functional impact on protein coding or regulatory DNA sequences, we performed whole-genome sequencing (WGS) of ER-positive T-47D human breast cancer cells co-expressing doxycycline-inducible A3B. and humanised bacteriophage PBS2 uracil glycosylase inhibitor (hUGI) peptide (**Figure 1A**, top panel) to inhibit the action of uracil-DNA glycosylase (UDG/UNG) and prevent excision of the A3B-catalysed uracil bases[22, 24]. The T-47D cell line was chosen because it does not express APOBEC3A which may confound the study[6], and also carries a loss-of-function *TP53* mutation that is known to protect the cells from the synthetic lethality caused by ectopic

expression of A3B[36]. After confirming the inducible expression of A3B/hUGI by immunoblotting, we exposed the cells to doxycycline for five days (**Figure S1A-B**), and sequenced genomic DNA samples from five individual T-47D clones (A+ colonies). As controls, sequencing results from three un-induced colonies (A- colonies) and three T-47D colonies (M+ colonies) co-expressing hUGI and enzymatically-dead A3B E68Q/E255Q double mutant (A3B**)[6] were used to represent single-nucleotide variants (SNV) arising either from the clonal separation process or base-editing by endogenous ABOBEC/AID-family enzymes (**Figure 1B**). Joint mutation calling by the GATK4 computational programme was carried out using a parental T-47D reference sample, the result of which showed a significantly elevated proportion of non-CpG C>T mutations in the A+ colonies over the M+ and A- colonies (**Figure 1C**). We next assessed the extent of focal hyper-mutation among the detected mutations using both intermutational distance[9] (**Figure 1D, S1C**) and an algorithm that detects somatic local hyper-mutation (HyperClust[37]) (**Figure 1E**). Unlike previous studies, where kataegis-like mutation clusters were prevalent in BER-functional models[32, 34], fewer clustered mutational events were detected in A+ colonies where BER activities were suppressed by hUGI. However, A+ colonies contained more non-recurrent and diffused form of clustered mutation termed omikli[37] (**Figure 1E**). Analysis of the resulting SNVs against pre-defined COSMIC single-base substitution (SBS) signatures (version 2) using non-negative matrix factorization (NMF)[30] identified the enrichment of signature SBS2 which was attributed to the activities of AID/APOBEC family of cytidine deaminases (**Figure 1F**). *De novo* NMF signature extraction revealed two prevalent signatures among the sequenced control colonies: signature A which is more prevalent in A-/M+ colonies and might account for the variation among T-47D subclones, and signature B which is mostly enriched in A+ colonies and harbours the A3B-preferred 3'-TCW trinucleotide motif, resembling the mutational process attributed to A3B editing (**Figure 1G, H**). Notably, other than C>T substitutions, a very weak contribution by C>A and C>G substitutions at the 5'-TC motif were observed in signature B, suggesting that most of the A3B-mediated uridine sites were subjected to BER processing. In addition, both

SBS2 and signature B are also found in M+ colonies, but enriched to a much lesser extent, implying the capture of intrinsic APOBEC/AID-family-mediated mutations by BER blockade. These data demonstrated that the BER-deficient breast cancer model successfully preserves and captures A3B activity on genomic DNA in the absence of the interference of downstream DNA repair mechanisms.

We next explored the functional implications of the captured A3B-editing sites using the Ensembl Variant Effect Predictor (VEP)[38]. When compared to random genomic locations, the TC>TT substitution in A+ and M+ colonies was significantly enriched in non-coding regulatory and transcription factor (TF) binding regions relative to protein-coding regions, an effect not seen in either VC>VT or ND>NN substitutions (**Figure 1I**). To further dissect function of these regulatory regions, we constructed an 18-state ChromHMM model for the T-47D cells using ChIP-seq data from six histone marks (**Figure S1D**) and computed its degree of overlap between the captured mutations (**Figure 1J**) using previously described methods[39]. In the A+ colonies we found overrepresentation of TC>TT mutations at transcription start sites (TSS) and enhancers, especially at the flanking regions of TSS and active enhancers. This pattern can also be found for intrinsic APOBEC-induced mutations in M+ colonies, but to a lesser extent than the A+ colonies. Analysis of DNA replication timing analysis using T-47D Repli-seq data from ENCODE indicated that the mapped TC>TT substitutions and omikli events in A+ colonies occurred early in DNA replication (**Figure S1E**). These findings are in accordance with previous observations that A3B targets genomic regions being actively transcribed and/or involved in early replication[37]. Overall, our data demonstrate a clear association between A3B-mediated base-editing and sites of gene transcription, consistent with a previously defined role for A3B in transcription regulation[6].

A3B binds and deaminates ssDNA in R-loops

We evaluated whether R-loops, a co-transcriptionally formed DNA:RNA hybrid, could provide substrates for A3B activity. In A+ colonies, TC>TT substitutions are flanked by a high level of GC-skewness for DNA sequences, which denotes overrepresentation of guanine bases in the

same DNA strand as the TC>TT editing event (**Figure 2A**). This also translates to an enrichment of G-quadruplex-forming sequences in ±500 bp regions around A3B-editing sites (**Figure 2A**). A similar effect can be seen around TC>TT substitutions from M+ colonies, albeit to a much lesser extent, but not around TC>TT or other di-nucleotide substitutions in the A-colonies (**Figure 2A, S2A-B**). Both GC-skew and G-quadraplex forming sequences are features reported to facilitate the formation or stabilisation of R-loops[40]. We mapped R-loops in parent T-47D cells using strand-specific DNA:RNA immunoprecipitation sequencing (ssDRIP-seq; **Figure S2C**) coupled with RNase H-dependency, GC-skewness and MNase-seq analyses to verify the identification of R-loops (**Figure S2D-E**). We found significant enrichment of R-loop signals in close proximity to TC>TT substitutions in A+ colonies. Strand-specific sequencing demonstrated that the TT>TC substitutions occurred on the ssDNA strand of the R-loop consistent with A3B preferentially editing the displaced ssDNA strand (**Figure 2B**). A weaker R-loop signal can be found at TC>TT substitutions from M+ colonies, but not in any other types of di-nucleotide substitutions (**Figure S2F**).

To further examine the association between R-loops and A3B-mediated editing, we first carried out co-immunoprecipitation experiments with the R-loop-binding S9.6 antibody, a previously reported methodology for isolating R-loops and their binding proteins[41]. Similar to DDX5, a previously validated R-loop binding protein[41], A3B co-complexed with immuno-precipitated R-loops, which was abrogated by the overexpression of RNaseH1, an enzyme that digests RNA:DNA hybrids (**Figure 2C**). Next, the genomic binding sites of flag-tagged A3B were mapped using chromatin immunoprecipitation sequencing (ChIP-seq) and ssDNA-associated protein immunoprecipitation (SPI-seq)[42] the latter of which is an approach better suited to capturing protein-bound ssDNA. More than half of A3B peaks detected by either method are found in the vicinity (< 1.5 kb) of R-loops. Although both methods yielded a similar proportion of R-loop-proximal A3B peaks (61% of all sequences with SPI-seq, 59% with standard ChIP-seq) the SPI-seq approach was more sensitive ($p < 0.01$, χ^2 test, **Figure 2D-E**). We also found that the SPI-seq captured more A3B-bound ssDNA fragments located in

the body of R-loops, in contrast to standard ChIP-seq which captures R-loop-flanking dsDNA (**Figure 2F**). Induced expression of RNaseH1 is known to reduce genomic R-loop content[7], and we found that A3B peaks adjacent to R-loops were specifically reduced by expressing RNaseH1 (**Figure 2E-G**), demonstrating a negative effect of R-loop clearance on A3B binding to genomic DNA. In addition, we observed a higher level of SPI-seq signal at TC>TT substitutions in A+ and M+ colonies compared with conventional ChIP-seq, and the binding of A3B to these sites was modulated by the expression of RNaseH1 (**Figure 2H-I, S2G**). This suggests R-loop formation facilitates A3B binding and cytidine deamination.

ER-induced R-loop formation facilitates A3B binding

Since activation of ER triggers the formation of co-transcriptional R-loops[7], we hypothesised that elevated levels of genomic R-loops would boost the availability of A3B substrate and in turn facilitate A3B binding to ER-activating enhancer regions. To test this, flag-tagged A3B binding sites and R-loops that are responsive to E2 stimulation were mapped using SPI-seq and ssDRIP-seq (**Figure 3A**) in T-47D breast cancer cells, which allowed the study of their spatial-temporal relationship. Similar to previous reports[6, 7], we observed an overall increase in A3B binding (23,711; 53% of all the peaks) and R-loop (18,567; 16% of all the peaks) peaks upon exposure to E2 for two hours (**Figure 3B**). Notably, following exposure to E2, a significant proportion of newly formed A3B binding sites are proximal to R-loops ($p < 10^{-5}$, χ^2 test), and correspondingly, R-loops near A3B binding sites are more susceptible to E2 induction (**Figure 3B**).

Because of the significant overlap between E2-induced A3B binding sites and R-loops (**Figure 3C**), we investigated a potential causal relationship between enhanced A3B binding and formation of R-loops. Depletion of A3B with siRNA does not affect the overall levels of R-loop signal or the increase in R-loops in response to E2-induction, as detected using slot blotting with S9.6 antibody (**Figure S3A-C**). Quantitative ssDRIP-seq also shows an increase of R-loops in proximity to A3B binding sites despite knockdown of A3B (**Figure S3D**). Using DESeq2 quantification method[43], we investigated the interaction between A3B siRNA and E2

treatments and observed minimal combinatorial effect on R-loop up-regulation between the two treatments (**Figure S3E**). More than half of E2-induced A3B peaks were situated either near E2-induced R-loops (34%) or R-loops that exist prior to E2 treatment (28%, **Figure 3D**), while R-loop-proximal A3B peaks had higher degree of induction upon E2 treatment relative to R-loop-distal A3B peaks (**Figure S3F**). These results indicate that elevated A3B binding is unlikely to be linked to induction of R-loops and instead that R-loop availability may determine E2-induced A3B binding and editing. To test this, we investigated whether overexpression of RNaseH1 affects A3B binding in five loci with well-defined ER binding and R-loop formation (**Figure 3E**). In contrast to A3B binding sites without R-loops, clearance of R-loops upon overexpression of RNaseH1 impairs A3B binding induced by exposure to E2 (**Figure 3F**). These data confirm that the binding of A3B to specific regions induced by exposure to E2 involves a subset of R-loops serving as substrates.

Our findings suggest a model where R-loop formation reshapes the genomic landscape around EREs to facilitate recruitment of A3B. Using our ChromHMM model, we found that E2-induced A3B binding sites that overlap with R-loops have higher enrichment scores in TSS, TSS flanking regions, and transcribed enhancers relative to those E2-induced A3B binding sites that are distal to R-loops (**Figure 3G**). Similarly, R-loops induced by E2 exposure are enriched in TSS and transcribed enhancers compared to the remaining pool of the non-induced R-loops, suggesting co-transcriptionally formed R-loops by E2 induction may facilitate A3B binding. Analysis of TF binding motifs reveals a different set of TFs potentially enriched in E2-regulated R-loop proximal peaks as opposed to R-loop-distal A3B peaks, where most sequence motifs in the former group are G-rich (**Figure S3G**). Collectively, our data suggest that enhanced R-loop availability induced by transactivation at ER enhancers or promoters may regulate A3B binding and editing activity.

A3B promotes DSB formation at E2-induced R-loop

Previous studies have shown that DSB formation induced by ER transactivation can be arise from the processing of either co-transcriptionally formed R-loops or A3B editing sites by DNA

damage repair mechanisms[6, 7]. To investigate the link between these processes, we employed DSBCapture-seq to precisely map genomic DSB sites[44], and to study the dependency of E2-induced DSBs on A3B and/or R-loops. We exposed T-47D breast cancer control and A3B-depleted cells to either DMSO or E2 for two hours, followed by DSBCapture-seq procedures (**Figure 4A**). Differential peak analysis of the sequencing results by DiffBind[45] revealed two groups of DSBs: DSBs that are induced by exposure to E2 ('E2-induced DSBs'), and DSBs modified by A3B knockdown ('A3B-modified DSBs', **Figure 4A**). Our results echo the reported widespread DNA damage following by ER activation[5-7, 46], and show that the 15,259 E2-induced DSB peaks accounts for 30% of all the detectable DSBs in the genome of the T-47D cell-line. Notably, 5,308 of the DSB peaks were modified by A3B-depletion, of which a majority (88%, $p < 10^5$ by X^2 test) overlapped with E2-induced DSBs (**Figure 4B**). This enrichment of A3B-modified DSBs in the E2-induced DSB set is consistent with previous reports[6]. Closer inspection of DSB sites identified in our study found that almost all (98.6%) of the A3B-modified DSB sites we detected overlapped with the E2-induced A3B binding sites identified by quantitative SPI-seq (**Figure 4C**), supporting a role for A3B in these DSBs. However, A3B binding alone is not sufficient to predict A3B-dependent formation of DSBs since 74% of the E2-induced A3B-independent DSBs, defined by up-regulated DSBs despite the depletion of A3B, also overlapped with A3B binding sites defined by SPI-seq. Instead, we found a requirement for R-loop formation prior to subsequent A3B binding to generate A3B-modified DSBs, since we detected a significant overlap with R-loops for these DSBs compared with E2-induced DSBs that were A3B-independent (**Figure 4C**). Accordingly, both R-loop formation and A3B binding distinguished A3B-modified DSBs from A3B-independent DSBs (96% vs 42%, **Figure 4D**) and were good indicators of the impact of A3B depletion on DSB formation (**Figure 4E**). Therefore, our results support the involvement of R-loop formation in the generation of A3B-induced DSBs.

A closer inspection of the regions flanking DSBs revealed some distinct features associated with A3B-dependent DSBs. These included higher bidirectional transcription activities (as

determined by GRO-seq), RNA polymerase II (PolII) occupancy, and denser G-quadruplex-forming sequences, all of which favour the formation of R-loops[47] (**Figure 4F**). Using our ChromHMM model, A3B-dependent DSBs were found to be more enriched in active/flanking TSS and transcribed enhancers, a group of chromatin states also enriched in E2-induced R-loops, suggesting an influence of R-loops on the DSB landscape similar to that seen for A3B binding sites (**Figure 4G**). These observations indicated a dependency on R-loops during the processing of A3B editing sites to DSBs. To test this, we analysed the five R-loop-overlapping A3B binding sites where A3B-dependent DSBs were found (**Figure 4H**) and tested the effect of RNAi-mediated depletion of Cockayne syndrome group B (CSB) protein. CSB is a TC-NER component known to be required for converting R-loops into DNA strand breaks[48] . Using quantitative PCR on DSBCapture-enriched DNA samples, we found that even modest CSB depletion was sufficient to reduce E2 responsive A3B-dependent DSB formation, but not E2-induced, A3B-independent DSBs (**Figure 4I**). These results revealed a requirement of R-loop processing into single-strand breaks (SSB) for the formation of A3B-dependent DSBs.

Blocking uracil base-processing impairs E2 response

The genomic location of A3B-binding sites, especially those in proximity to R-loops and linked to the formation of DSBs, suggested a gene regulatory role for the down-stream processing of A3B base-editing. To address this, we performed RNA-seq studies on T-47D breast cancer cells harbouring a doxycycline inducible enzymatically inactive A3B** and hUGI cassette (**Figure 5A**). This construct prevents the downstream processing of the intrinsic A3B-mediated C>U editing sites by inhibiting UNG enzymes. We collected RNA samples after exposing cells to E2 for 16 hours, with or without doxycycline induction, and performed RNA-seq. For differential gene expression analysis, interaction modelling by DEseq2 programme was employed to dissect the impact of UNG inhibition (by hUGI) over E2 response[43]. Results showed that the expression of A3B**-hUGI dampened the E2-regulated response of 87% of genes (**Figure 5B**), out of which 90 genes were identified as statistically significant with an interaction model by DESeq2. (**Figure 5C**). Using pathway enrichment analysis with MSigDB

hallmark gene set[49], we observed a significant overlap between these genes and genes regulated by estrogen response pathways (**Figure 5D**), indicating a requirement for uracil repair during the transactivation of ER. In order to investigate whether the genes likely to be regulated by A3B are affected by lack of uracil-repair mechanisms, we first used the predictive algorithm rGREAT to generate two groups of genes which are most likely to associated with regions that could form gene-regulating DSBs[50], the first one contained 86 genes that were associated with A3B-dependent DSBs, and the latter one contained 333 genes that were associated with all E2-induced A3B binding sites that were proximal to R-loops. Using our RNA-seq data as the input for leading edge analysis by gene set enrichment analysis (GSEA)[51], we found a significant overlap between genes from both of these groups and with genes up-regulated by E2 (**Figure 5E**), suggesting the formation of A3B-dependent DSBs may facilitate the transactivation of ER-regulated genes. More importantly, the genes from both groups were also enriched with E2-regulated genes that were negatively affected by A3B**-hUGI expression which confers BER-deficiency (**Figure 5E**), suggesting that the inability to process A3B-editing sites into DSBs in certain regulatory regions hampered the transactivation of gene expression following E2 stimulation. Together, our collective results demonstrate that the downstream processing of A3B C>U editing sites, which leads to formation of DSBs, may promote E2 response in T-47D cells.

TDRD3 guides A3B binding to regulatory elements

We found that ER activation leads to pervasive formation of R-loops with potential to be A3B substrates; however, only a subset of these R-loops was bound by A3B, suggesting a molecular mechanism that facilitates R-loop selectivity. To test whether this potential selectivity could be driven by proteins interacting with A3B, we employed BioID[52], a proximity-dependent labelling method, to capture A3B binding partners. A3B tagged with BirA* on either N- or C-terminal was used as bait, and BirA*-red fluorescent protein (RFP) was used as control. In addition to T-47D cells, the BioID experiments were also carried out in two additional cell lines, MDA-MB-468 with high A3B expression background, and MCF-10-A with very low

A3B levels (**Figure 6A, Figure S4A**). *Bona-fide* A3B interactors were selected against common BioID backgrounds using SAINTexpress score in conjunction with historical BioID non-specific controls curated in the CRAPome database[53]. Our results showed a general overlap of A3B interactors across different cell lines, which also found a higher rate of detection in higher intrinsic A3B expressing cancer cells than the low-expression MCF-10-A cells (**Figure 6B**). Using A3B interactors detected in at least two cell lines, we plotted an interaction network using common protein-protein interaction databases (**Figure 6C**) and found two clusters of interactions: one with mainly RNA-binding or HNRNP components and the other comprising members of the TOP3B-TDRD3-FMRP (TTF) complex, a protein complex known in function in R-loop resolution within cells[54]. To validate the interactions between the TTF complex and A3B, we carried out co-immunoprecipitations and found A3B was able to interact with TDRD3 and FMRP either with or without the bridging effect of nucleic acids (**Figure 6D**). Upon closer inspection, we found that A3B binds only isoform 1 of TDRD3, but not the shorter isoforms 2 and 3, which lack part of an N-terminal OB-fold, suggesting the OB-fold site may contribute to A3B binding. TOP3B was not detected in the A3B complex, despite evidence for binding to TDRD3 at the N-terminus domain (**Fig. 6D**). As TOP3B also scored as a hit in the A3B BioID, there may be a dynamic but potentially mutually exclusive binding interaction between A3B, TOP3B and TDRD3.

To further investigate the link between TDRD3 and the binding of A3B to DNA, we conducted TDRD3 ChIP-seq in T-47D cells as well as A3B SPI-seq in cells depleted of TDRD3 by RNAi. Knockdown of TDRD3 attenuated approximately 78% of A3B binding sites identified by SPI-seq when compared to control cells, suggesting that A3B binding requires TDRD3 (**Figure 6E**). Furthermore, analysis of ChIP-seq data showed that these putative TDRD3-dependent A3B sites also displayed a strong TDRD3 binding capability (**Figure 6F-G**). We performed ChIP-seq experiments using an antibody (Millipore $H3^{R17me2a}$) that recognises a spectrum of asymmetric dimethylarginine (ADMA)-modified proteins which are validated TDRD3 interactors[55], and found these proteins co-localised with A3B binding sites identified by SPI-

seq (**Figure 6F-G**). These results suggest that ADMA-modified proteins bind A3B and regulate its binding to DNA. When compared to DRIP-seq and DSBCapture results, a majority (62%) of the putative TDRD3-dependent A3B binding sites are R-loop-proximal, and notably, these sites also tend to overlap A3B-dependent DSBs (**Figure 6F**). From the perspective of TDRD3 binding sites, a large proportion (79%) of the *bona fide* binding sites identified using ChIP-seq overlapped with A3B binding sites, the signals of which were also attenuated by TDRD3 knockdown (**Figure S4B**). Also, more than half of the A3B-overlapping TDRD3 binding sites were proximal to R-loops, showing a bias towards A3B-dependent DSBs, and had a similar ChromHMM profile to A3B-dependent DSBs (**Figure S4B-C**). Detailed quantitative studies on seven putative TDRD3-dependent A3B binding sites (**Figure 6H**) using DRIP-qPCR and SPI-qPCR on T-47D cells confirmed a requirement of intrinsic TDRD3 on A3B binding, while depletion of TDRD3 has minimal impact on R-loop formation on these sites (**Figure 6I-J**). Therefore, our results suggests that TDRD3 may play an important role to guide the binding of A3B to regulatory elements and may serve as a mechanism to regulate its function in converting R-loops into DSBs during regulation of transcription.

Taken together, our results support a mechanistic model (**Figure 7**) in which TDRD3 functions as an adapter protein directing the DNA deamination activity of A3B to ssDNA strand of R-loops upon transactivation of ER, which also induces ADMA post-translational modifications of protein complexes localising to enhancers or promoters. The induction of co-transcriptional R-loops and A3B editing upon ER activation leads to ER-induced DNA DSBs, which modulate the chromatin configuration of enhancer or promoter regions and contribute to chromosomal instability in ER-driven cancers.

Discussion

Thus far, numerous endeavours have been attempted to detect, sample and measure A3B-driven DNA deamination sites, advancing our understanding of A3B function in cancer[10, 30, 32-34]. However, direct capturing and mapping of A3B sites in cancer cells has so far proven to be

challenging, mainly because of the masking effect of the highly effective BER and a TP53-dependent synthetic lethal interaction between BER-deficiency and A3B enzymatic activity[31, 36]. Here, we circumvented this problem by using a BER-deficient, A3B-overexpressing human breast cancer cell line model in a *TP53* functional loss context. By sequencing a number of clones, our approach captured A3B-driven mutations in bulk, as supported by strand-coordinated focal TC>TT hypermutations and NMF signatures with known A3B traits. The strong association between these sites with regulatory elements (TSS and enhancers), transcriptional activity and early DNA replication timing strengthens the role of A3B as a TF (in our case, ER) co-activator. We encourage additional studies beyond ER-driven models to examine whether A3B could also have a similar effect on other TFs, especially other nuclear receptors.

Using the A3B deamination sites detected by WGS, we also added R-loops into the known repository of A3B substrates in addition to lagging DNA synthesis strand[33, 35, 56], loop regions of DNA hairpins[10, 57], and ssDNA intermediates during recombination and repair[9, 10, 58]. The predominant pathway processing A3B-catalysed uracil lesions in the context of R-loops is the BER[24], which has been observed in previous studies in APOBEC-Cas9-mediated R-loop editing and the AID-catalysed deamination within R-loops during immunoglobulin class switch recombination[27, 28]. However, if R-loops are resolved before repair of the uracil lesion, the re-annealing of the template strand will lead to a U:G mismatch which can also be repaired by DNA mismatch repair (MMR). Unlike the BER that faithfully repairs the U:G lesion, MMR is known to cause omikli mutations and is detectable using datasets from TCGA project and ICGC consortium[59, 60]. Nevertheless, DNA sequences flanking putative A3B-driven omikli mutations detected in these real-world datasets did not display any R-loop traits, suggesting that A3B-catalysed uracil bases in R-loops were predominantly repaired by BER.

Importantly, we also demonstrated that the repair of A3B-catalysed uracil bases in R-loops is part of a mechanism leading to the formation of R-loop-mediated DSBs, which are known to affect chromosome integrity and contribute to genomic heterogeneity in cancer cells[47, 61]. It

has been shown previously that R-loop clearance by TC-NER factors, such as XPF/XPG enzymes, removes stretches of DNA/RNA hybrid and results in SSBs[48]. Further collision with the replication fork/transcriptional machinery, or actions by unidentified structure-specific endonucleases are required to generate DSBs[47, 48, 61]. Here, we have identified a unique way of R-loop-mediated DSB formation, which does not require the above-mentioned prerequisites but is achieved instead by the introduction of SSBs by BER-processing of A3B-deaminated bases at the opposite strand of R-loop-induced SSBs. This mechanism further details the role of A3B in inducing DSBs at ER binding sites[6], and our observations unify the roles of R-loops and A3B during the formation of DSB following ER activation[6, 7].

A fuller understanding of how SSBs effect R-loop homeostasis will require further investigation. Recently, McCann et al. demonstrated that a high-level expression of A3B in cells results in an accelerated clearance rate of steady-state- or stimulation-induced-R-loops[29]. We speculate that this process would also be accompanied by an elevated level of DNA DSBs as many of the A3B-dependent DSBs display a requirement for R-loop formation in our study. More sophisticated studies will be needed to determine whether A3B-mediated SSBs occur prior to the clearance of DNA:RNA hybrid, and if so, whether R-loops with a SSB would serve as a better substrate for TC-NER.

Using BioID-based mass-spectroscopy proteomics, we identified TDRD3 as a binding partner leading to the recruitment of A3B to enhancer/TSS R-loops. TDRD3 is an oncogene and a transcriptional coactivator[62, 63], and can be recruited to active chromatin by the binding of its Tudor domain to ADMA-modified proteins[64]. TDRD3 exerts its function by acting as an adapter, linking these ADMA-modified proteins to effector proteins, such as TOP3B and DHX9 that resolve R-loops[54, 65]. Despite both being R-loop processors, A3B may have distinct functions compared with TOP3B, that may lead to different transcriptional outcomes. This hypothesis is supported by our observation that A3B and TOP3B compete for the same binding site on TDRD3. Previous studies have shown that loss of either TOP3B or A3B could result in R-loop accumulation and elevated DNA damage[29]. Therefore, it will be interesting to

compare the DSB-inducing capabilities of these two enzymes in future studies and identify mechanisms leading to the selection of TDRD3 partners.

Our results also demonstrate that R-loop selection by A3B may be achieved by binding of TDRD3 to ADMA-modified proteins including histones ($H3^{R17me2a}$/$H4^{R3me2a}$), Pol II (Pol $II^{R1810me2a}$) and the mediator complex subunit 12 ($MED12^{R1899me2a}$)[66]. These ADMA-modified proteins mark promoter, enhancer, or looping regions, which harbour bi-directional transcriptional activity and could form R-loops with long intergenic noncoding RNA (lincRNA) or enhancer RNA (eRNA) binding[67]. Using hUGI as a surrogate of functional-loss of A3B and transcriptional profiling by RNA-seq, we showed that ER-mediated transactivation programmes rely on A3B-mediated DNA editing in these regions, consolidating A3B as a regulator of ER activity independent of ER-ligand associations.

Previous reports have proposed A3B as a therapeutic target in a model where elevated A3B drives tamoxifen-resistance through the acquisition of increased A3B-dependent gene mutations. Our data also support the therapeutic targeting of A3B in tamoxifen-resistant breast cancer but with an emphasis on its role as a driver of an adaptive response at the transcriptional or epigenetic level rather than through direct gene mutation. Our findings also support the development of A3B inhibitors as new therapies to control ER activity in the context of selective ER modulator-resistant cancers and, potentially, additional nuclear hormone receptor-driven malignancies.

Materials and methods

Cell culturing and cell line preparation

T-47D (ATCC HTB-133), MCF-10A (ATCC CRL-10317) and MDA-MB-468 (HTB-132) cells were obtained from ATCC. Lenti-X 293T cells were obtained from Clonetech. Cells were maintained using ATCC recommended medium except Lenti-X 293T cells were cultured with DMEM medium supplemented with 10% v/v FBS (PAA) and 0.5% v/v 10000U/ml penicillin /streptomycin solution. All cells were cultured at 37 ºC with 5% CO_2.

To generate stable inducible cell lines, T-47D human breast cancer cells at exponential grow phase were transfected with lentivirus at a M.O.I of 100 with the assist of 10 μg/ml polybrene. After a transduction period of 48 hours, selection was carried out using medium supplemented with 4 μg/mL puromycin (Gibco). Transduction efficiency was monitored by either turbo-RFP expression, A3B immuno-fluorescence microscopy or immunoblotting following induction of sample cells with doxycycline. After verification, lentivirus cassettes were maintained by culturing cells with 4 ug/mL puromycin.

Drug treatments

For the induction of lentiviral protein expression, cells were pre-exposed to full culture medium supplemented with 1 μg/mL doxycycline (Sigma) for 24 hours unless denoted otherwise. Follow-up drug treatments, if included, also contained the same concentration of doxycycline throughout the entire treatment period.

For ER induction by E2 (Selleck Chem), adherent cells were briefly washed by warm PBS twice before incubated with phenol-red-free RPMI-1640 medium (Gibco) supplemented with 10% charcoal-stripped FBS (Thermo Fisher) for 48 hours at 37 ºC. E2 was added to the culture medium at a final concentration of 100 nM, and further incubated for two hours (unless denoted otherwise) before the cells were harvested for follow-up studies.

Colony separation and WGS

0.25 million exponentially growing T-47D cells harbouring lentiviral 3×Flag-A3B-P2A-hUGI-HA or 3×Flag-A3B**-P2A-hUGI-HA were seeded onto a 100 mm plate for 24 hours, followed by induction with 1 μg/ml doxycycline for 120 hours. Cells were harvested using trypsin digestion, serial diluted and plated onto a 96-well plate (Corning). Colonies derived from single cells were picked and the cloned cells were expanded with culture medium supplemented with 4 μg/ml puromycin. Genomic DNA was extracted from each colony using a Qiagen Genomic-tips 500/G kit. Sequencing libraries for whole genome sequencing with DNA nanoball (DNB) technology were constructed (Beijing Genomics Institute (BGI) Inc., Hong Kong). Sequencing

was carried out using BGISEQ-500 sequencer with a mean sequencing coverage of greater than 30× for each of the samples using 2 × 150 bp configuration.

Further method details are provided in the supplementary text.

Data availability

WGS, ChIP-seq, SPI-seq, ssDRIP-seq, DSBCapture-seq and RNA-seq datasets reported in this study are available as dataset GSE193234 in the Gene Expression Omnibus (GEO) database. Processed results for RNA-seq are available in Supplemental Table 1, and processed BioID mass spectrometry are available in Supplemental Table 2. Previously reported datasets used in this study are listed in the supplemental text.

Acknowledgements

This work is financially supported by Cancer Research UK (C309/A11566 and C2739/A22897) and ICR (London, United Kingdom). P.W. acknowledges additional grant support from the Wellcome Trust (212969/Z/18/Z), Cancer Research UK (C35696/A23187) and The Mark Foundation/The Chordoma Foundation. C.Z. was sponsored by Shanghai Pujiang Programme (22PJD104). We thank Prof. Ping Yuan from Sun Yat-sen University, Dr. Mike Walton and Dr. Alexandra Vasile from ICR for helpful discussions.

Author Contributions

C.Z., O.W.R., P.A.C, and P.W. conceived and designed this study. C.Z., Y.L., M.W., and M.T. conducted the experiments and analysis unless otherwise noted. B.C.,Z.B., K.M. and B.A-L conducted analysis on sequencing data. provided guidance on data analysis using sequencing data. C.Z., A.H., M.P.L, O.W.R., P.A.C, and P.W. prepared the manuscript.

Competing Interests Statement

C.Z., A.H., M.P.L., O.W.R., P.A.C, K.M., M.T., and P.W. are current or past employees of The Institute of Cancer Research, which has a commercial interest in the discovery and

development of A3B inhibitors. P.W. is an independent director at Storm Therapeutics, is a consultant/advisory board member at Astex Pharmaceuticals, CV6 Therapeutics, Black Diamond Therapeutics, Vividion Therapeutics and Nextechinvest; reports receiving a commercial research grant from Sixth Element Capital, Astex Pharmaceuticals, and Merck; has ownership interest in Storm Therapeutics, Chroma Therapeutics, and Nextechinvest; and has an unpaid consultant/advisory board relationship with the Chemical Probes Portal. P.W. has also received relevant research funding from Vernalis and Astex Pharmaceuticals. No potential conflicts of interest were disclosed by the other authors.

References:

1 Orzolek I, Sobieraj J, Domagala-Kulawik J. Estrogens, Cancer and Immunity. *Cancers (Basel)* 2022; **14**.
2 Arao Y, Korach KS. The physiological role of estrogen receptor functional domains. *Essays Biochem* 2021; **65**:867-875.
3 Yang JA, Stires H, Belden WJ, Roepke TA. The Arcuate Estrogen-Regulated Transcriptome: Estrogen Response Element-Dependent and -Independent Signaling of ERalpha in Female Mice. *Endocrinology* 2017; **158**:612-626.
4 Hewitt SC, Korach KS. Estrogen Receptors: New Directions in the New Millennium. *Endocr Rev* 2018; **39**:664-675.
5 Rajan A, Nadhan R, Latha NR, Krishnan N, Warrier AV, Srinivas P. Deregulated estrogen receptor signaling and DNA damage response in breast tumorigenesis. *Biochim Biophys Acta Rev Cancer* 2021; **1875**:188482.
6 Periyasamy M, Patel H, Lai CF *et al.* APOBEC3B-Mediated Cytidine Deamination Is Required for Estrogen Receptor Action in Breast Cancer. *Cell Rep* 2015; **13**:108-121.
7 Stork CT, Bocek M, Crossley MP *et al.* Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. *Elife* 2016; **5**.
8 Li JJ, Weroha SJ, Lingle WL, Papa D, Salisbury JL, Li SA. Estrogen mediates Aurora-A overexpression, centrosome amplification, chromosomal instability, and breast cancer in female ACI rats. *Proc Natl Acad Sci U S A* 2004; **101**:18123-18128.
9 Nik-Zainal S, Alexandrov LB, Wedge DC *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* 2012; **149**:979-993.
10 Nik-Zainal S, Davies H, Staaf J *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 2016; **534**:47-54.
11 Compe E, Egly JM. TFIIH: when transcription met DNA repair. *Nat Rev Mol Cell Biol* 2012; **13**:343-354.
12 Kamileri I, Karakasilioti I, Garinis GA. Nucleotide excision repair: new tricks with old bricks. *Trends Genet* 2012; **28**:566-573.
13 Fong YW, Cattoglio C, Tjian R. The intertwined roles of transcription and repair proteins. *Mol Cell* 2013; **52**:291-302.
14 Schuermann D, Weber AR, Schar P. Active DNA demethylation by DNA repair: Facts and uncertainties. *DNA Repair (Amst)* 2016; **44**:92-102.
15 Perillo B, Castoria G, Migliaccio A. Exploiting the mechanism of estrogen-induced transcription to fight breast cancer. *Exp Mol Med* 2021; **53**:1205-1206.
16 Puc J, Kozbial P, Li W *et al.* Ligand-dependent enhancer activation regulated by topoisomerase-I activity. *Cell* 2015; **160**:367-380.
17 Ju BG, Lunyak VV, Perissi V *et al.* A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription. *Science* 2006; **312**:1798-1802.
18 Swanton C, McGranahan N, Starrett GJ, Harris RS. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer Discov* 2015; **5**:704-712.

19 Green AM, Weitzman MD. The spectrum of APOBEC3 activity: From anti-viral agents to anti-cancer opportunities. *DNA Repair (Amst)* 2019; **83**:102700.
20 Harris RS, Dudley JP. APOBECs and virus restriction. *Virology* 2015; **479-480**:131-145.
21 Henderson S, Fenton T. APOBEC3 genes: retroviral restriction factors to cancer drivers. *Trends Mol Med* 2015; **21**:274-284.
22 Burns MB, Lackey L, Carpenter MA *et al.* APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* 2013; **494**:366-370.
23 Burns MB, Temiz NA, Harris RS. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat Genet* 2013; **45**:977-983.
24 Petljak M, Dananberg A, Chu K *et al.* Mechanisms of APOBEC3 mutagenesis in human cancer cells. *Nature* 2022; **607**:799-807.
25 Law EK, Sieuwerts AM, LaPara K *et al.* The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. *Sci Adv* 2016; **2**:e1601737.
26 Garcia-Muse T, Aguilera A. R Loops: From Physiological to Pathological Roles. *Cell* 2019; **179**:604-618.
27 Gu S, Bodai Z, Cowan QT, Komor AC. Base Editors: Expanding the Types of DNA Damage Products Harnessed for Genome Editing. *Gene Genome Ed* 2021; **1**.
28 Pilzecker B, Jacobs H. Mutating for Good: DNA Damage Responses During Somatic Hypermutation. *Front Immunol* 2019; **10**:438.
29 McCann JL, Cristini A, Law EK *et al.* R-loop homeostasis and cancer mutagenesis promoted by the DNA cytosine deaminase APOBEC3B. *bioRxiv* 2021.
30 Alexandrov LB, Kim J, Haradhvala NJ *et al.* The repertoire of mutational signatures in human cancer. *Nature* 2020; **578**:94-101.
31 Krokan HE, Saetrom P, Aas PA, Pettersen HS, Kavli B, Slupphaug G. Error-free versus mutagenic processing of genomic uracil--relevance to cancer. *DNA Repair (Amst)* 2014; **19**:38-47.
32 Akre MK, Starrett GJ, Quist JS *et al.* Mutation Processes in 293-Based Clones Overexpressing the DNA Cytosine Deaminase APOBEC3B. *PLoS One* 2016; **11**:e0155391.
33 Hoopes JI, Cortez LM, Mertz TM, Malc EP, Mieczkowski PA, Roberts SA. APOBEC3A and APOBEC3B Preferentially Deaminate the Lagging Strand Template during DNA Replication. *Cell Rep* 2016; **14**:1273-1282.
34 Nikkila J, Kumar R, Campbell J *et al.* Elevated APOBEC3B expression drives a kataegic-like mutation signature and replication stress-related therapeutic vulnerabilities in p53-defective cells. *Br J Cancer* 2017; **117**:113-123.
35 Sui Y, Qi L, Zhang K *et al.* Analysis of APOBEC-induced mutations in yeast strains with low levels of replicative DNA polymerases. *Proc Natl Acad Sci U S A* 2020; **117**:9440-9450.
36 Serebrenik AA, Starrett GJ, Leenen S *et al.* The deaminase APOBEC3B triggers the death of cells lacking uracil DNA glycosylase. *Proc Natl Acad Sci U S A* 2019; **116**:22158-22163.
37 Mas-Ponte D, Supek F. DNA mismatch repair promotes APOBEC3-mediated diffuse hypermutation in human cancers. *Nat Genet* 2020; **52**:958-968.
38 McLaren W, Gil L, Hunt SE *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* 2016; **17**:122.
39 Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* 2017; **12**:2478-2492.
40 Lee CY, McNerney C, Ma K, Zhao W, Wang A, Myong S. R-loop induced G-quadruplex in non-template promotes transcription by successive R-loop formation. *Nat Commun* 2020; **11**:3392.
41 Cristini A, Groh M, Kristiansen MS, Gromak N. RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. *Cell Rep* 2018; **23**:1891-1905.
42 Zhou ZX, Zhang MJ, Peng X *et al.* Mapping genomic hotspots of DNA damage by a single-strand-DNA-compatible and strand-specific ChIP-seq method. *Genome Res* 2013; **23**:705-715.
43 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; **15**:550.
44 Lensing SV, Marsico G, Hansel-Hertsch R, Lam EY, Tannahill D, Balasubramanian S. DSBCapture: in situ capture and sequencing of DNA breaks. *Nat Methods* 2016; **13**:855-857.
45 Rory S, Gord, Brown. Differential Binding Analysis of ChIP-Seq peak data. https://github.com/aeron15/DiffBind 2015.
46 Williamson LM, Lees-Miller SP. Estrogen receptor alpha-mediated transcription induces cell cycle-dependent DNA double-strand breaks. *Carcinogenesis* 2011; **32**:279-285.
47 Aguilera A, Garcia-Muse T. R loops: from transcription byproducts to threats to genome stability. *Mol Cell* 2012; **46**:115-124.

48 Sollier J, Stork CT, Garcia-Rubio ML, Paulsen RD, Aguilera A, Cimprich KA. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol Cell* 2014; **56**:777-785.
49 Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 2015; **1**:417-425.
50 McLean CY, Bristor D, Hiller M *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 2010; **28**:495-501.
51 Subramanian A, Tamayo P, Mootha VK *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005; **102**:15545-15550.
52 Roux KJ, Kim DI, Burke B, May DG. BioID: A Screen for Protein-Protein Interactions. *Curr Protoc Protein Sci* 2018; **91**:19 23 11-19 23 15.
53 Mellacheruvu D, Wright Z, Couzens AL *et al.* The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods* 2013; **10**:730-736.
54 Yang Y, McBride KM, Hensley S, Lu Y, Chedin F, Bedford MT. Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. *Mol Cell* 2014; **53**:484-497.
55 Cheng D, Vemulapalli V, Lu Y *et al.* CARM1 methylates MED12 to regulate its RNA-binding ability. *Life Sci Alliance* 2018; **1**:e201800117.
56 Seplyarskiy VB, Soldatov RA, Popadin KY, Antonarakis SE, Bazykin GA, Nikolaev SI. APOBEC-induced mutations in human cancers are strongly enriched on the lagging DNA strand during replication. *Genome Res* 2016; **26**:174-182.
57 Buisson R, Langenbucher A, Bowen D *et al.* Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features. *Science* 2019; **364**.
58 Taylor BJ, Nik-Zainal S, Wu YL *et al.* DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *Elife* 2013; **2**:e00534.
59 Cancer Genome Atlas Research N, Weinstein JN, Collisson EA *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 2013; **45**:1113-1120.
60 Hutter C, Zenklusen JC. The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. *Cell* 2018; **173**:283-285.
61 Sollier J, Cimprich KA. Breaking bad: R-loops and genome integrity. *Trends Cell Biol* 2015; **25**:514-522.
62 Morettin A, Paris G, Bouzid Y *et al.* Tudor Domain Containing Protein 3 Promotes Tumorigenesis and Invasive Capacity of Breast Cancer Cells. *Sci Rep* 2017; **7**:5153.
63 Narayanan N, Wang Z, Li L, Yang Y. Arginine methylation of USP9X promotes its interaction with TDRD3 and its anti-apoptotic activities in breast cancer cells. *Cell Discov* 2017; **3**:16048.
64 Yang Y, Lu Y, Espejo A *et al.* TDRD3 is an effector molecule for arginine-methylated histone marks. *Mol Cell* 2010; **40**:1016-1023.
65 Yuan W, Al-Hadid Q, Wang Z *et al.* TDRD3 promotes DHX9 chromatin recruitment and R-loop resolution. *Nucleic Acids Res* 2021; **49**:8573-8591.
66 Suresh S, Huard S, Dubois T. CARM1/PRMT4: Making Its Mark beyond Its Function as a Transcriptional Coactivator. *Trends Cell Biol* 2021; **31**:402-417.
67 Tan-Wong SM, Dhir S, Proudfoot NJ. R-Loops Promote Antisense Transcription across the Mammalian Genome. *Mol Cell* 2019; **76**:600-616 e606.

Figure and legends:

Figure 1: Detection of A3B deamination sites in BER-deficient cell models.

(A). Lentiviral inducible system used in this study and immunoblotting of cells with and without 24-hour induction by 1 μg/ml doxycycline. A3B** denotes A3B with E68Q/E225Q mutations.

(B). Schematic of sample preparations for WGS studies.

(C). Mean relative contribution of the indicated types of point mutations for the three indicated sample groups. Error bars = SD; *: $p < 0.05$; ****: $p < 10^{-4}$, one-way ANOVA.

(D). Representative waterfall plot of intermutational distance (IMD) of each mutation identified in the three indicated groups of colonies. Dotted line denotes IMD ≤ 10^3 bp.

(E). Mutation clusters detected by HyperClust for three sample groups. Data for mean value and error bars for SD. *: $p < 0.05$; **: $p < 0.01$, one-way ANOVA.

(F). Heat map showing the cosine similarity scores of top-ranking COSMIC signatures for each indicated samples.

(G). Relative contribution of each indicated trinucleotide alternation to the two top-ranking *de novo* mutational signatures identified by NMF analysis.

(H). Heat map showing the cosine similarity scores of mutational signatures in G for each indicated samples.

(I). Distribution of variant consequence as determined by VEP for indicated types of variants. †: random sampling of 10^6 sites. ****: $p < 10^{-4}$, χ^2 test comparing to random sites.

(J). Heat map showing enrichment score of chromatin states for ± 50 bp regions flanking point mutations. Chromatin states were derived by ChromHMM (**Figure S1D**).

Figure 2: A3B binds and edits R-loop.

(A). Profiles of GC skewness and frequency of G-quadruplex (G4) motifs in regions flanking the TC>TT SNP. Line denotes average value and shaded areas 95% CI.

(B). Heat maps showing ssDRIP-seq signals in regions flanking the TC>TT SNP identified in A+ colonies (left), their corresponding signal profile plots (middle), and schematics demonstrating the position of R-loop in relation to the site of TC>TT SNPs (right).

(C). S9.6 co-immunoprecipitation assay of T-47D cells transfected with or without RNaseH1-encoding vectors for 24 hours (left), and representative immunoblots and slot blots from input samples (left).

(D). Venn-diagram showing the overlap of Flag-A3B binding sites from ChIP-seq and SPI-seq. Numbers of sites proximal to (within 1.5 kb) or distal to R-loop are shown.

(E). Heat maps of enrichment signal from indicated sequencing experiments at the flanking regions of consensus Flag-A3B ChIP-seq and SPI-seq peaks.

(F). Profile of Flag-A3B ChIP-seq or SPI-seq signal for R-loops that are proximal to A3B sites. Line denotes average value and shaded areas 95% CI. Statistic represents one-way ANOVA using enrichment scores at indicated regions.

(G). Tukey boxplots showing enrichment scores from SPI-seq experiments. ****: $p \leq 10^{-4}$, two-way ANOVA assessing the effect of R-loop proximity on response to 24-hour RNaseH1 overexpression.

(H). Heat maps showing Flag-A3B ChIP-seq or SPI-seq signal in regions flanking TC>TT SNP identified in A+ colonies.

(I). Signal profile for Flag-A3B ChIP-seq or SPI-seq signal in regions flanking TC>TT SNP identified in A+ colonies. Line denotes average value and shaded areas 95% CI.

Figure 3: R-loops induced by ER activation facilitates A3B binding.

(A). Schematic of the quantitative A3B SPI-seq and experiment carried in this study.

(B). Volcano plot summarising the effect of two-hour E2 stimulation on Flag-A3B binding sites determined by SPI-seq (left) and R-loop sites determined by ssDRIP-seq (right). R-loop-proximal Flag-A3B sites are coloured in red. Up-regulated sites are defined as

FDR ≤ 0.05 with fold change ≥ 1.5 (indicated by red dotted line). P value represent statistical significance by χ^2 test of non-random association between R-loop or A3B proximity and response to E2.

(C). Venn diagram showing the overlap between E2-induced Flag-A3B binding sites and E2-induced R-loops. P value was obtained by χ^2 test for the independence of the two sets in comparison with the genomic background.

(D). Heat maps showing Flag-A3B SPI-seq, ssDRIP-seq and ESR1 ChIP-seq signals in ±8 kb regions flanking E2-induced Flag-A3B binding sites.

(E). Representative signal tracks for ESR1 ChIP-seq, Flag-A3B SPI-seq and ssDRIP-seq at E2-induced Flag-A3B binding sites that are proximal to E2-induced R-loops (CISH, PDPR, PGR and RARA) or R-loops pre-existing to E2-induction (AGR3 and CABLES1).

(F). Bar graph showing Flag-A3B SPI-qPCR and DRIP-qPCR results. Data represent mean of three biological replicates and error bars for SD. *: p ≤ 0.05; **: p ≤ 0.01; ***: p ≤ 0.001; ns: p > 0.05, two-way ANOVA assessing the effect size of RNaseH1 treatment on E2 response.

(G). Heat map showing enrichment scores for ChromHMM chromatin states at indicated A3B binding sites or R-loops.

Figure 4: A3B promotes E2-induced DSB formation at R-loops.

(A). Schematic of quantitative DSBCapture experiments to identify E2-induced and A3B-dependent DSBs in T-47D cells.

(B). Pie chart showing results of quantitative DSBCapture experiments, p value derived from χ^2 test evaluating the statistical significance of non-random association between A3B-dependent and E2-induced DSBs.

(C). Heat maps visualising normalised signal for identified DSBs in T-47D cells. Each row represents one DSB site and each column a unique sample, and types of overlapping A3B or R-loop features are displayed alongside.

(D). Bar graph showing percentage of DSB peaks overlapping A3B and/or R-loop.

(E). Tukey boxplots showing log_2 ratio of DSBCapture signals between A3B and non-targeted siRNA-transfected cells after 2 hours of E2 transfection. Statistical analysis denotes two-way ANOVA, ****: $p \leq 10^{-4}$.

(F). Heat maps for DSBCapture, A3B SPI-seq, ssDRIP-seq and ESR1 ChIP-seq signals flanking ±8 kb of centre of DSBs. Also shown are bars indicating types of overlapping Flag-A3B SPI-seq and ssDRIP-seq peaks.

(G). Profiles of GRO-seq, PolII ChIP-seq signal and G-quadruplex occurrences in regions flanking indicated types of DSBs. Line represents average signal or frequency and shaded area denote 95% C.I.

(H). Heat map showing enrichment scores for ChromHMM chromatin states at indicated DSBs.

(I). Representative signal tracks for DSBCapture at E2-induced Flag-A3B binding sites that are proximal to E2-induced R-loops.

(J). Bar graph showing DSBCapture-qPCR results. T-47D cells were treated with NT or CSB siRNA for 48 hours before 2-hour stimulation by DMSO or 100 nM E2. Data represent mean of three biological replicates and error bars for SD. *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$; two-way ANOVA assessing the effect size of CSB siRNA on E2 response.

Figure 5: Blocking base excision repair impairs E2 response.

(A). Schematic of the RNA-seq experiment in this study.

(B). Dot plot showing log_2 fold change of normalised transcript counts for 819 differentially expressed genes affected by E2 stimulation without the expression of A3B**-hUGI.

(C). Heat maps showing normalised RNA-seq signals for 90 differentially expressed genes the E2 response of which were modulated by A3B**-hUGI induction. Statistical results from DESeq2 are summarised in bars on the right.

(D). Table showing MSigDB hallmark gene set enrichment analysis using 90 hit genes described in (C).

(E). Leading edge analysis for genes associated with A3B-dependent DSB or R-loop-overlapping E2-induced Flag-A3B binding sites. Associations of genomic regions and gene were predicted using rGREAT using a binomial Bonferroni-adjusted p value of 0.05. Leading edge analysis were conducted using fGSEA package with R.

Figure 6: TDRD3 facilitates A3B binding to regulatory elements.

(A). Bubble plot illustrating the result of BioID experiments. T-47D cells bearing lentiviral inducible BioID baits were induced with doxycycline for 24 hours before subject to BioID analysis. SAINT express FDR score were derived using BirA*-RFP pull-down results from T-47D cells as control and calculated by CRAPome. Data derived from n = 3 independent experiments and recurring hits were labelled.

(B). Venn-diagram showing the overlap between BioID hits (FDR < 0.05) across three different breast cancer cell lines.

(C). Network diagram showing known protein-protein interactions amongst hits that qualified in at least two different cell lines.

(D). Representative immunoblotting using antibodies against A3B or TDRD3 in T-47D cells.

(E). Venn-diagram showing the overlap between Flag-A3B binding sites in NT siRNA-treated or TDRD3 siRNA-treated T-47D cells, which were identified by SPI-seq. Immunoblots verifying the knockdown of TDRD3 are shown at the bottom.

(F). Heat map illustrating signals from indicated sequencing experiments in ±8 kb regions flanking Flag-A3B binding sites identified by SPI-seq. Also shown are bars indicating types of overlapping DSBCapture peaks. Statistic results are from χ^2 test evaluating the statistical significance of non-random association between overlapped R-loops and A3B-modified DSBs. **** and ns denote $p \leq 10^{-4}$ and $p > 0.05$ respectively.

(G). Profiles of TDRD3, $H3^{R17me2a}$ and CARM1 in regions flanking A3B peak centres identified by SPI-seq.

(H). Representative signal tracks for TDRD3 ChIP-seq and A3B SPI-seq at indicated genomic loci.

(I). Bar graphs showing results from DRIP-qPCR experiments. Samples were collected from T-47D cells transfected with NT siRNA, TDRD3 siRNA or vector encoding

RNaseH1 for 48 hours. Data represent mean of n = 3 biological replicates and error bars for SD. *, ** and *** denote p < 0.05, 0.01 and 0.001 respectively as measured by one-way ANOVA comparing to siNT control.

(J). Bar graphs showing results for A3B SPI-qPCR experiments. Chromatins were collected from T-47D cells transfected with indicated siRNA for 48 hours. Data represent n = 3 biological replicates and error bars for SD. * and ** denote p < 0.05 and 0.01 respectively as measured by two-tailed t-test.

Figure 7: Mechanistic model detailing the generation of DSBs as a result of collaborative processing of R-loops and R-loop-editing by A3B, guided by TDRD3 adapter protein upon ER transactivation.

Figure 1

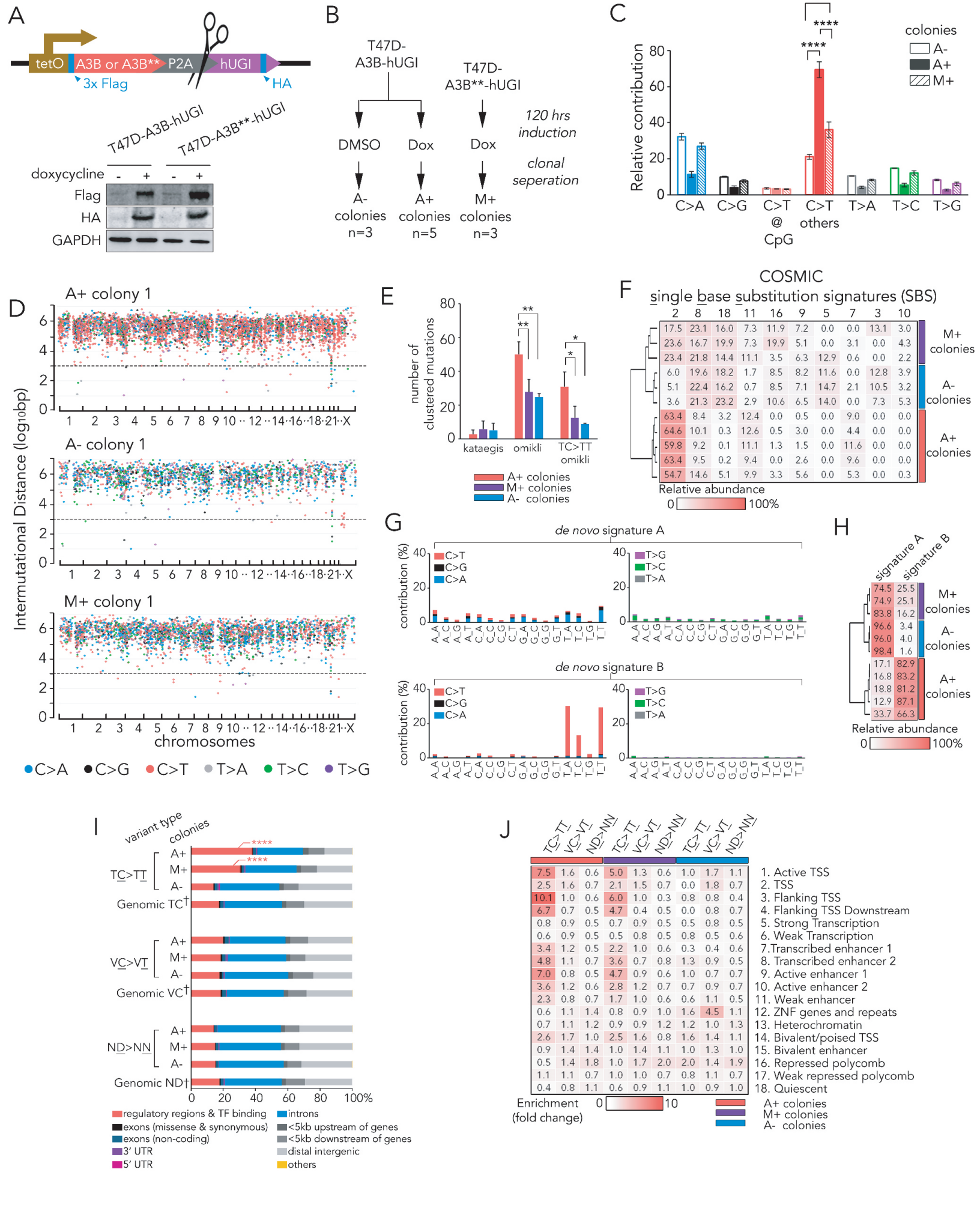


Figure 2

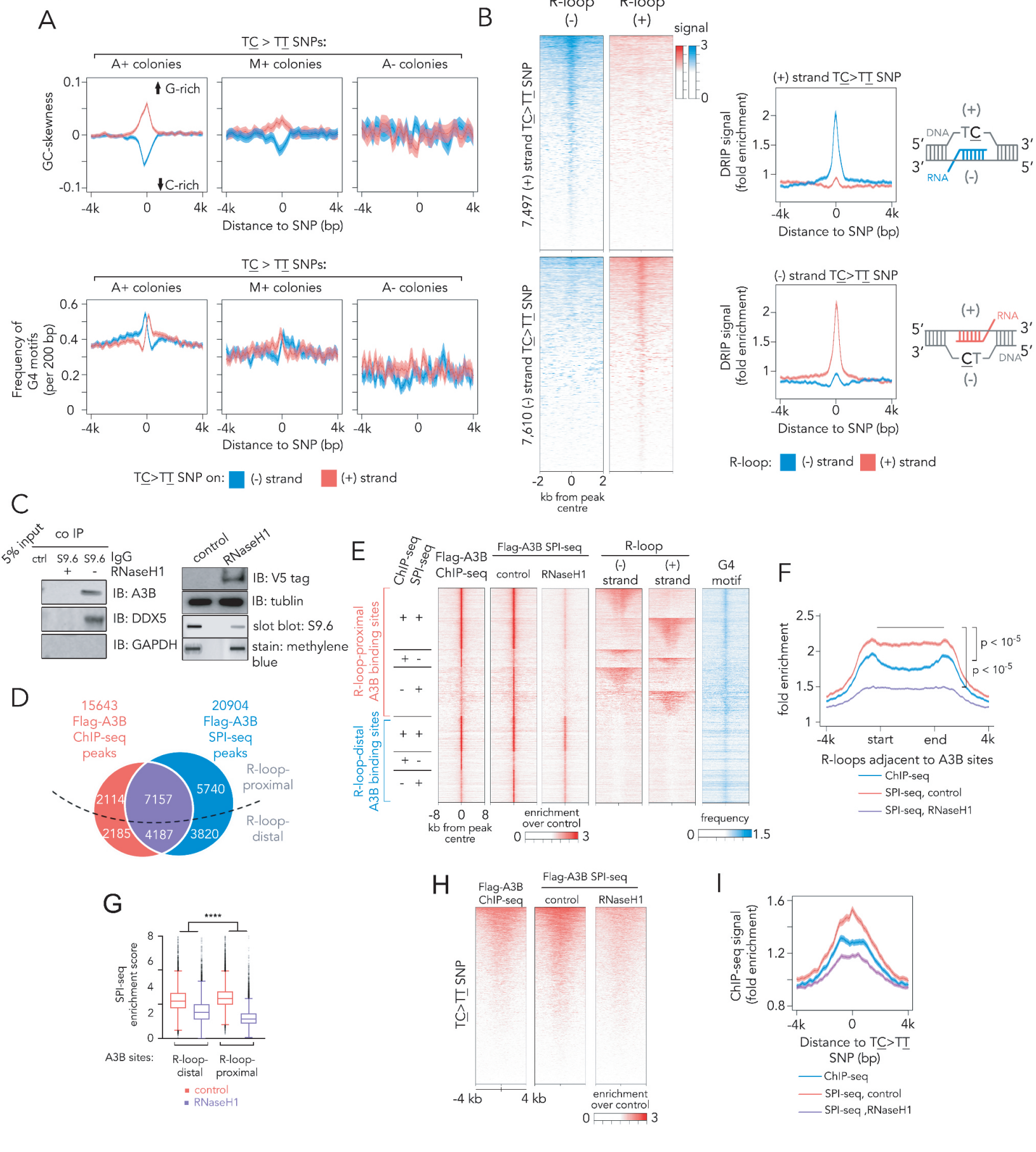


Figure 3

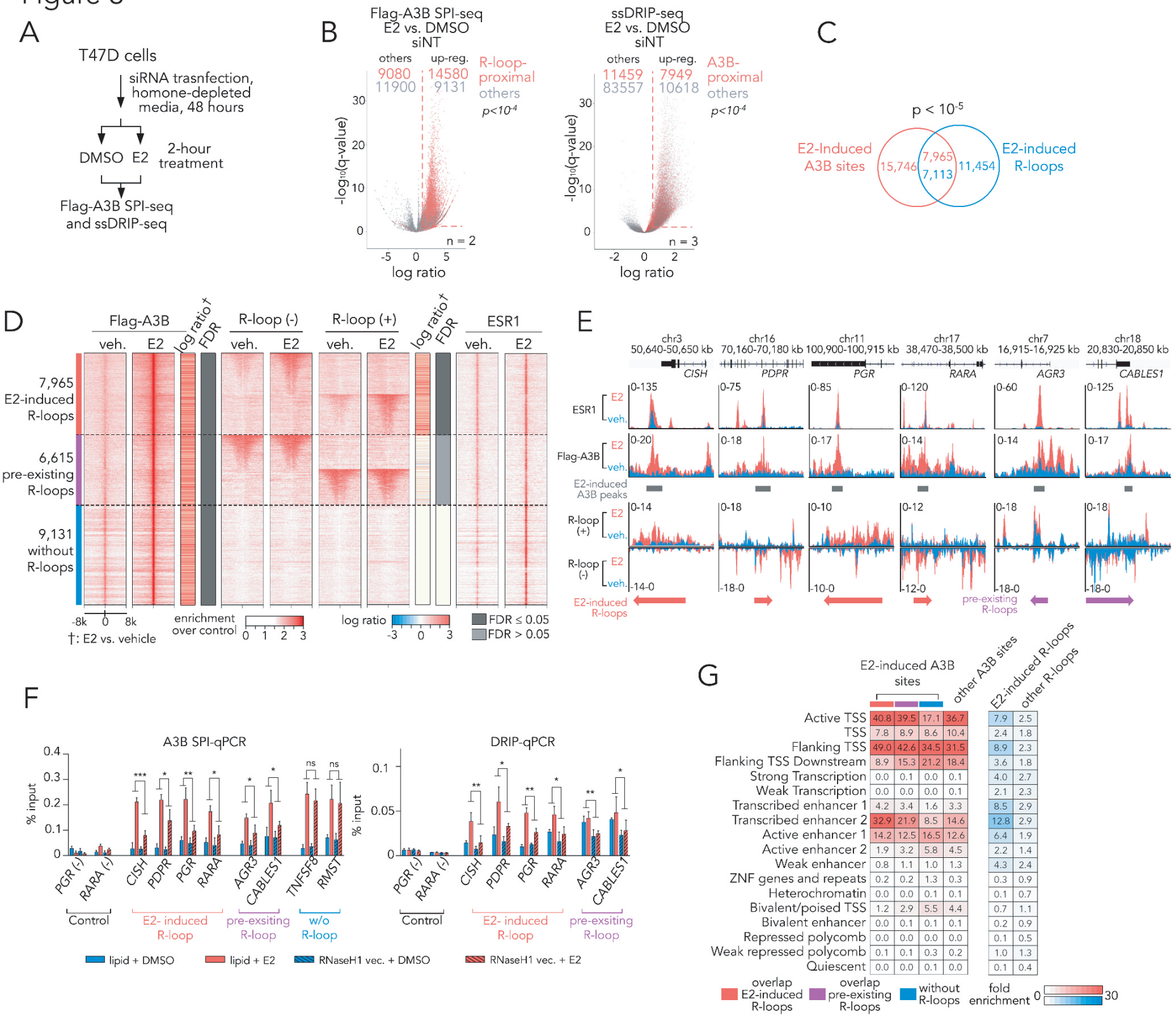

A
T47D cells
siRNA trasnfection, homone-depleted media, 48 hours
DMSO E2
2-hour treatment
Flag-A3B SPI-seq and ssDRIP-seq
B
Flag-A3B SPI-seq
E2 vs. DMSO
siNT
others
up-reg.
9080
11900
14580
9131
R-loop-proximal
others
p<10-4
-log10(q-value)
n = 2
log ratio
ssDRIP-seq
E2 vs. DMSO
siNT
11459
83557
7949
10618
A3B-proximal
n = 3
C
p < 10-5
E2-Induced A3B sites
15,746
7,965
7,113
11,454
E2-induced R-loops
D
Flag-A3B
R-loop (-)
R-loop (+)
ESR1
veh.
E2
log ratio†
FDR
24,156 E2-induced A3B binding sites
7,965 E2-induced R-loops
6,615 pre-existing R-loops
9,131 without R-loops
-8k 0 8k
†: E2 vs. vehicle
enrichment over control
log ratio
FDR ≤ 0.05
FDR > 0.05
E
chr3 50,640-50,650 kb
chr16 70,160-70,180 kb
chr11 100,900-100,915 kb
chr17 38,470-38,500 kb
chr7 16,915-16,925 kb
chr18 20,830-20,850 kb
CISH
PDPR
PGR
RARA
AGR3
CABLES1
ESR1
Flag-A3B
E2-induced A3B peaks
R-loop (+)
R-loop (-)
E2-induced R-loops
pre-existing R-loops
F
A3B SPI-qPCR
DRIP-qPCR
% input
PGR (-)
RARA (-)
CISH
PDPR
PGR
RARA
AGR3
CABLES1
TNFSF8
RMST
Control
E2- induced R-loop
pre-exsiting R-loop
w/o R-loop
lipid + DMSO
lipid + E2
RNaseH1 vec. + DMSO
RNaseH1 vec. + E2
G
E2-induced A3B sites
other A3B sites
E2-induced R-loops
other R-loops
Active TSS 40.8 39.5 17.1 36.7 7.9 2.5
TSS 7.8 8.9 8.6 10.4 2.4 1.8
Flanking TSS 49.0 42.6 34.5 31.5 8.9 2.3
Flanking TSS Downstream 8.9 15.3 21.2 18.4 3.6 1.8
Strong Transcription 0.0 0.1 0.0 0.1 4.0 2.7
Weak Transcription 0.0 0.0 0.0 0.1 2.1 2.3
Transcribed enhancer 1 4.2 3.4 1.6 3.3 8.5 2.9
Transcribed enhancer 2 32.9 21.9 8.5 14.6 12.8 2.9
Active enhancer 1 14.2 12.5 16.5 12.6 6.4 1.9
Active enhancer 2 1.9 3.2 5.8 4.5 2.2 1.4
Weak enhancer 0.8 1.1 1.0 1.3 4.3 2.4
ZNF genes and repeats 0.2 0.2 1.3 0.3 0.3 0.9
Heterochromatin 0.0 0.0 0.1 0.1 0.1 0.7
Bivalent/poised TSS 1.2 2.9 5.5 4.4 0.7 1.1
Bivalent enhancer 0.0 0.1 0.1 0.1 0.2 0.9
Repressed polycomb 0.0 0.0 0.0 0.0 0.1 0.5
Weak repressed polycomb 0.1 0.1 0.3 0.2 1.0 1.3
Quiescent 0.0 0.0 0.1 0.0 0.1 0.4
overlap E2-induced R-loops
overlap pre-existing R-loops
without R-loops
fold enrichment
0
30

Figure 4

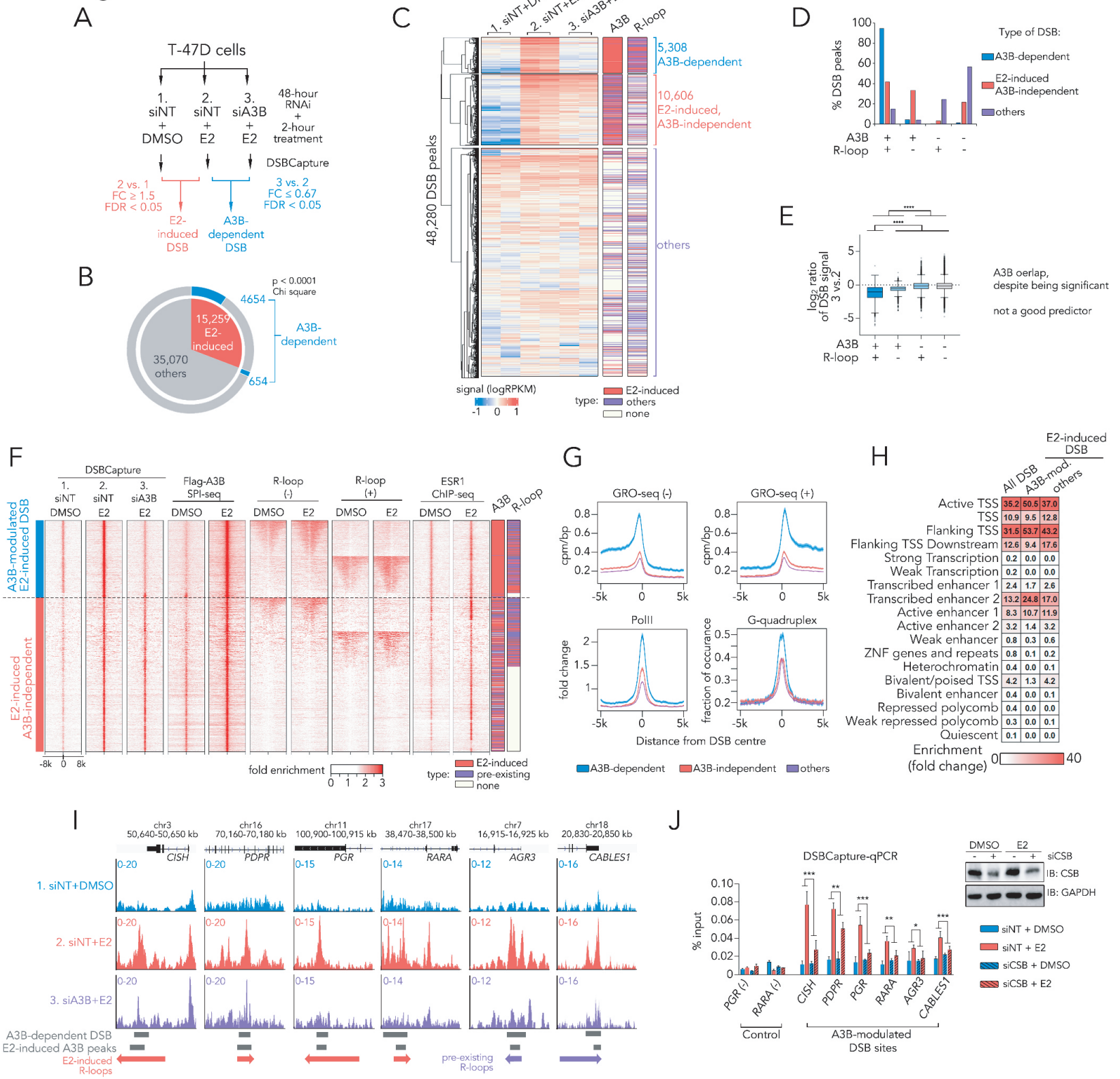
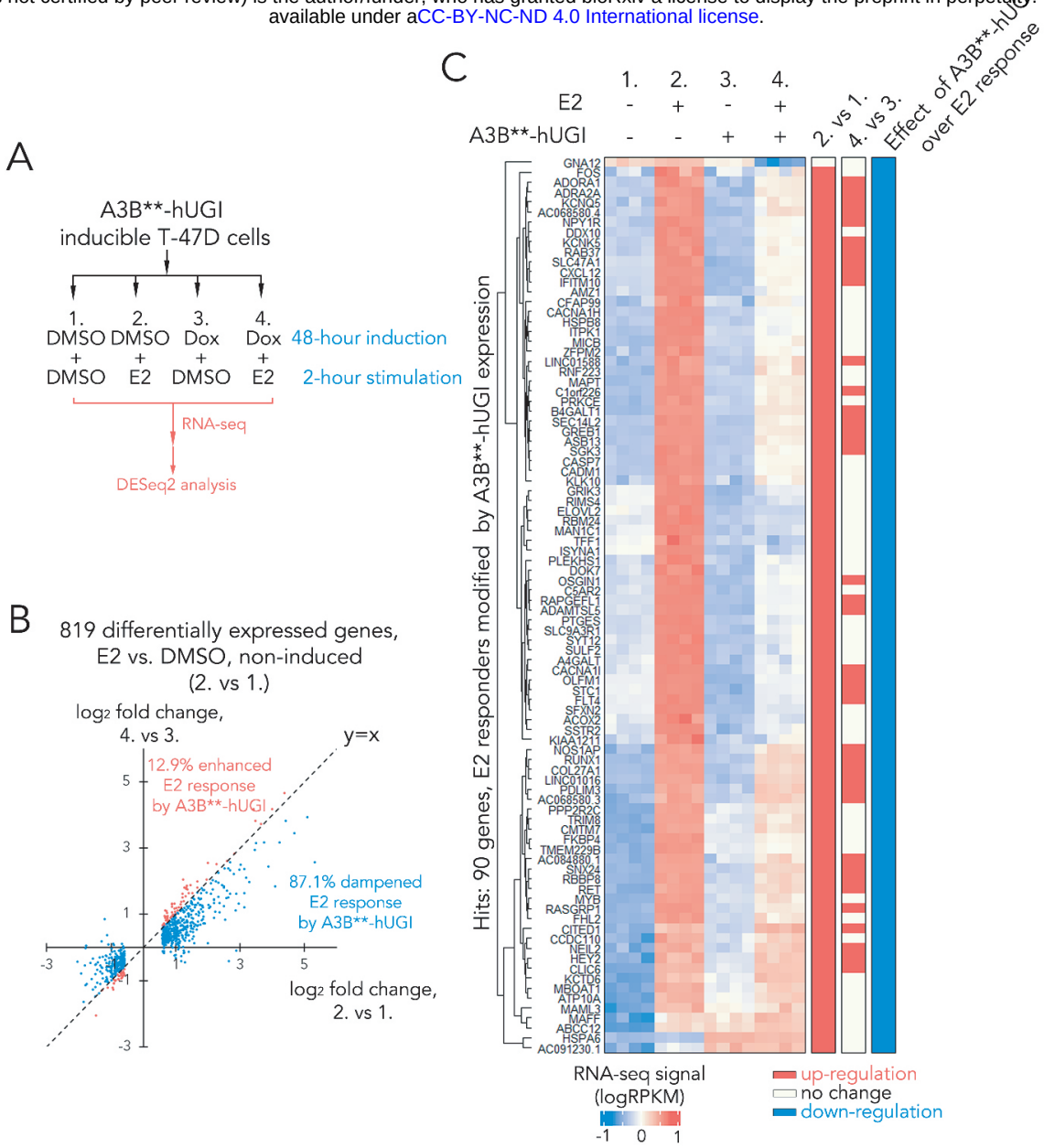

A
T-47D cells
1. siNT + DMSO
2. siNT + E2
3. siA3B + E2
48-hour RNAi + 2-hour treatment
DSBCapture
2 vs. 1 FC ≥ 1.5 FDR < 0.05
3 vs. 2 FC ≤ 0.67 FDR < 0.05
E2-induced DSB
A3B-dependent DSB
B
p < 0.0001 Chi square
4654
15,259 E2-induced
35,070 others
654
A3B-dependent
C
1. siNT+DMSO
2. siNT+E2
3. siA3B+E2
A3B
R-loop
5,308 A3B-dependent
10,606 E2-induced, A3B-independent
others
48,280 DSB peaks
signal (logRPKM)
type: E2-induced others none
D
% DSB peaks
Type of DSB:
A3B-dependent
E2-induced A3B-independent
others
E
log2 ratio of DSB signal 3 vs.2
A3B oerlap, despite being significant
not a good predictor
F
DSBCapture
Flag-A3B SPI-seq
R-loop (-)
R-loop (+)
ESR1 ChIP-seq
A3B-modulated E2-induced DSB
E2-induced A3B-independent
fold enrichment
type: E2-induced pre-existing none
G
GRO-seq (-)
GRO-seq (+)
PolII
G-quadruplex
cpm/bp
fold change
fraction of occurance
Distance from DSB centre
A3B-dependent A3B-independent others
H
E2-induced DSB
All DSB
A3B-mod.
others
Active TSS 35.2 50.5 37.0
TSS 10.9 9.5 12.8
Flanking TSS 31.5 53.7 43.2
Flanking TSS Downstream 12.6 9.4 17.6
Strong Transcription 0.2 0.0 0.0
Weak Transcription 0.2 0.0 0.0
Transcribed enhancer 1 2.4 1.7 2.6
Transcribed enhancer 2 13.2 24.8 17.0
Active enhancer 1 8.3 10.7 11.9
Active enhancer 2 3.2 1.4 3.2
Weak enhancer 0.8 0.3 0.6
ZNF genes and repeats 0.8 0.1 0.2
Heterochromatin 0.4 0.0 0.1
Bivalent/poised TSS 4.2 1.3 4.2
Bivalent enhancer 0.4 0.0 0.1
Repressed polycomb 0.4 0.0 0.0
Weak repressed polycomb 0.3 0.0 0.1
Quiescent 0.1 0.0 0.0
Enrichment (fold change)
I
chr3 50,640-50,650 kb CISH
chr16 70,160-70,180 kb PDPR
chr11 100,900-100,915 kb PGR
chr17 38,470-38,500 kb RARA
chr7 16,915-16,925 kb AGR3
chr18 20,830-20,850 kb CABLES1
1. siNT+DMSO
2. siNT+E2
3. siA3B+E2
A3B-dependent DSB
E2-induced A3B peaks
E2-induced R-loops
pre-existing R-loops
J
DSBCapture-qPCR
% input
PGR (-)
RARA (-)
CISH
PDPR
PGR
RARA
AGR3
CABLES1
Control
A3B-modulated DSB sites
DMSO
E2
siCSB
IB: CSB
IB: GAPDH
siNT + DMSO
siNT + E2
siCSB + DMSO
siCSB + E2

Figure 5



D

Enrichment analysis using 90 E2-responders modified by A3B**-hUGI expression

Hallmark gene set	hit/total	FDR
Estrogen response early	28/200	4.00E-42
Estrogen response late	20/200	8.42E-27
UV response (up)	6/158	1.86E-05
Glycolysis	4/200	8.70E-03

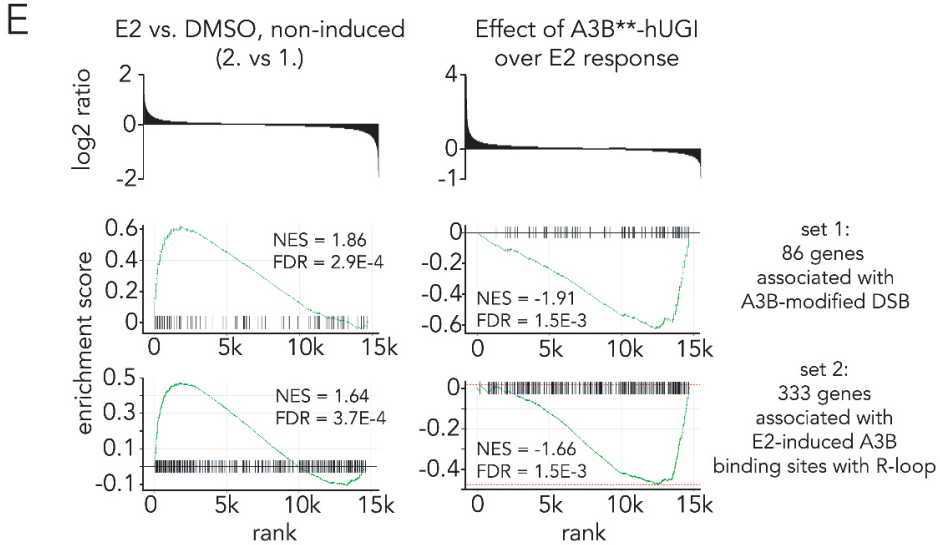


Figure 6

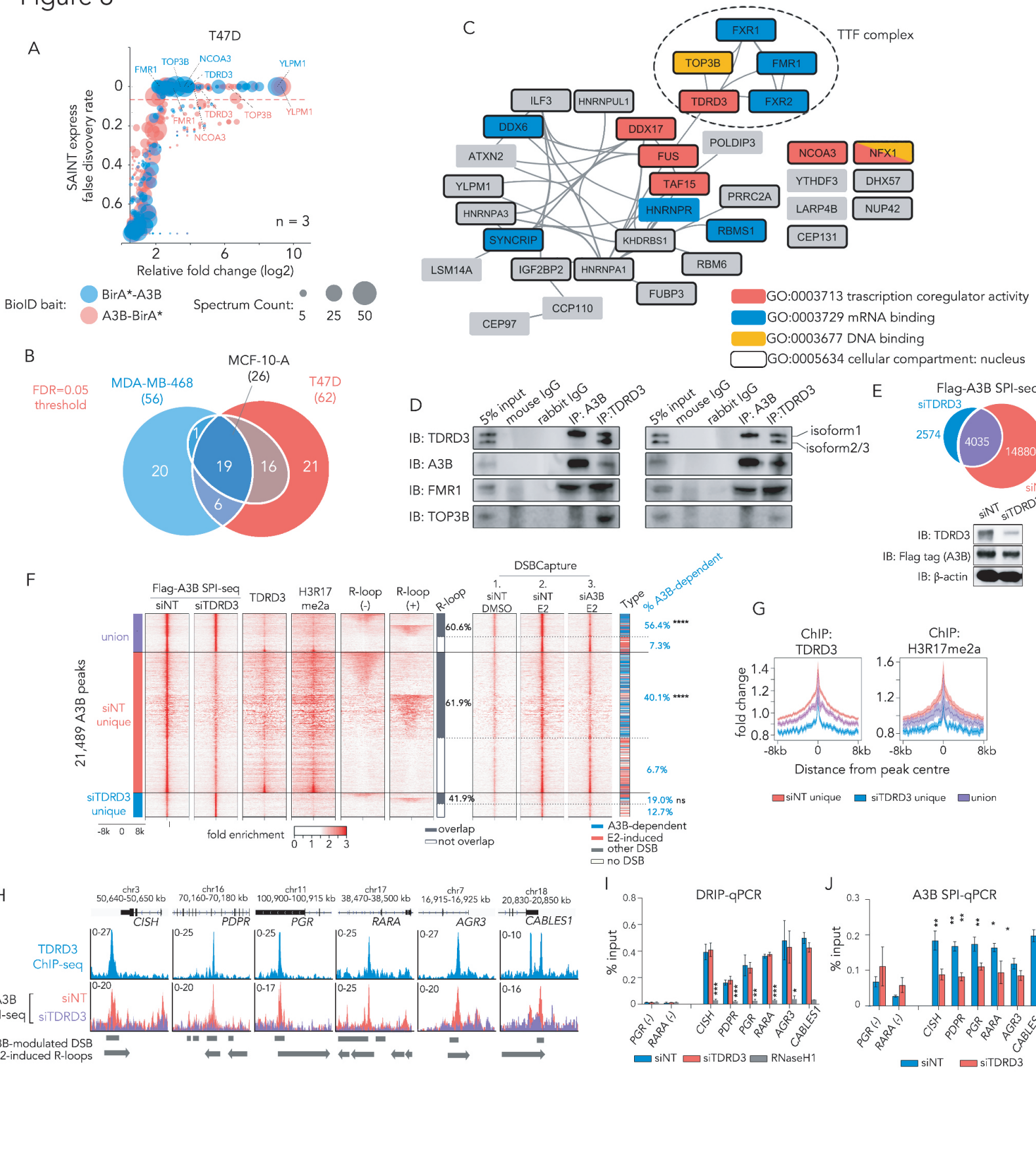


Figure 7