Super-enhancers shape the landscape and repair dynamics of transcription-associated DNA breaks in cancer

Osama Hidmi[1], Diala Shatleh[1], Sara Oster Flayshman[1], Jonathan Monin[1], and Rami I. Aqeilan[1,2,]

[1]The Concern Foundation Laboratories, The Lautenberg Center for Immunology and Cancer Research, Department of Immunology and Cancer Research-IMRIC, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel
[2]Lead contact

ramiaq@mail.huji.ac.il (R.I.A.)

Abstract

Cancer is characterized by uncontrolled proliferation accompanied by the hypertranscription of oncogenes, leading to transcription stress, a key source of DNA double-strand breaks (DSBs) that jeopardize genomic stability. Yet, transcription stress is still underexplored. In this study, we utilized maps of DSBs identified through in-suspension break labeling *in situ* and sequencing (sBLISS), along with transcription stress markers, revealing that transcription stress regions coincide with the super-enhancer regulatory landscape. Notably, γH2AX mapping indicates its enrichment at transcription stress sites, while not all DSB-enriched genes show equal γH2AX marking, but those with DSBs tied to transcription stress are distinctly marked. Intriguingly, genes with high-DSBs marked by γH2AX exhibited significantly higher DSB turnover and repair than those with γH2AX-low genes, manifesting vulnerability to mutagenesis. These findings underscore super-enhancer activity as a determinant of the transcription stress landscape in cancer, posing a threat to the genomic stability of oncogenes.

Introduction

The most fundamental hallmark of cancer cells is their ability to sustain proliferation, which they acquire through genetic alterations during neoplastic transformation[1]. Simultaneously, dysregulation of the transcriptional program emerges to accommodate the cell's increased demand for cellular factors essential for proliferation, growth, and survival[2,3]. Consequently, genes encoding these factors, i.e., oncogenes, are hypertranscribed to fulfill the needs of cancer cells[4].

A well-established mechanism by which cancer cells upregulate oncogenes involves the formation of oncogenic super-enhancers (SEs)[5–12]. In normal cells, SEs drive the expression of cell-type-specific genes that encode master transcription factors (MTFs), which shape cellular identity and the transcriptional program[2]. In cancer cells, SEs drive the expression of cancer-type-specific oncogenes and oncogenic MTFs that dictate their transcriptional programs and support their proliferation[2,12]. SEs can become dysregulated in cancer through various mechanisms: 1) overexpression of enhancer transcription factors renders SEs much more effective at driving the expression of their associated genes; 2) translocations can relocate SEs near an inactive oncogene in the corresponding normal cell, and 3) focal amplifications of SEs can lead to the increased expression of target genes[12]. These mechanisms contribute to the activation of proto-oncogenes by SEs, which drive the transcription of these genes to an exceptionally high level termed "hypertranscription."

Transcription is increasingly viewed as a source of DNA damage[13–18]. The most damaging type of DNA lesions produced during transcription is DNA double-strand breaks (DSBs)[19]. These DSBs often result indirectly from the processing of transcription-associated intermediates, such as single-strand breaks (SSBs) that form during the repair of trapped topoisomerase 1 cleavage complexes (TOP1cc). DSBs resulting from transcription are primarily facilitated by the formation of pathogenic co-transcriptional R-loops and the aberrant activity of topoisomerase 1 (TOP1)[13,20]. R-loops tend to form more frequently at highly expressed genes, and their formation causes TOP1 to become trapped in its cleavage complex form (TOP1cc)[13].

Although transcription is a highly regulated process, it can occasionally become abortive and result in DSBs. These transcriptional aberrations occur more frequently as transcription levels increase, in a phenomenon known as "transcription stress"[21]. Characterizing the "hotspots" of transcriptional stress is crucial because transcription-related DSBs can lead to cell death,

genomic instability, and mutations[22,23]. Moreover, since cancer cells upregulate genes vital for their proliferation and survival, a deeper understanding of transcriptional stress and transcription-associated DSBs associated with these oncogenes can aid in exploiting their vulnerabilities to target cancer cells more effectively. Nevertheless, while transcription stress has been increasingly studied[24–26], its landscape and consequences, particularly in cancer, remain incompletely characterized.

DSBs are impediments to transcription[27,28] and must, therefore, be efficiently repaired to sustain continuous high transcriptional levels. However, it is not understood how regions of transcription stress are equipped to handle these deleterious DSBs.

In this study, we characterize the landscape of transcription stress in cancer cells and reveal how these regions are structured to promote the rapid and efficient repair of transcriptional DSBs that sustain hypertranscription.

Results

Super-enhancer-regulated oncogenes are major transcription stress sites.

Pathogenic R-loops preferentially form at highly expressed genes[29], facilitating aberrant TOP1 activity and TOP1cc trapping[13]. The high TOP1 occupancy, TOP1cc trapping, and co-transcriptional R-loops all contribute to the formation of transcription-associated DSBs [20,30]. Therefore, we reasoned that these features, along with genome-wide maps of endogenous DSBs (hereafter referred to as "Breakome"), could serve as markers of transcriptional stress. To determine transcription stress sites in the genome, we analyzed sBLISS and ChIP-seq data for the breast cancer cell line MCF7 and applied a Hidden Markov Model (HMM)-based approach to identify regions in the genome with co-occurrence across the four datasets (sBLISS, DRIP-seq, TOP1cc, and TOP1 ChIP-seq). As expected, the majority of these regions were genic (Extended Data Fig. 1A), while the rest were enriched mainly at intergenic super-enhancers (SEs) (Extended Data Fig. 1B and C). To investigate which genes serve as transcription stress sites, we scored each gene based on TOP1, TOP1cc, R-loops, and DSBs scores (see Methods). This scoring yielded a shortlist of genes, many of which have well-established oncogenic functions (Fig. 1A). Upon closer analysis of this list, we noted that many of the genes were regulated by super-enhancers (SEs) in MCF7 cells, including CCND1, NEAT1, miR21, and MYC (SE-regulated genes were downloaded from SEdb 2.0[31], see methods) (Fig. 1A). To statistically

validate this observation, we systematically assessed the representation of multiple gene sets among transcription stress sites by performing gene set enrichment analysis (Fig 1B; see Methods). SE-regulated genes were the most significantly enriched category, with SE-regulated oncogenes exhibiting the highest enrichment (~122-fold over expectation under a random model, $P = 1.85 \times 10^{-15}$; hypergeometric test). Similar results were obtained when quantifying the overlap between each gene set and the co-marked regions identified by HMM (Extended Data Fig. 1D-E). Notably, the enrichment of SE-regulated genes was substantially higher than that of typical-enhancer (TE)-regulated genes or highly expressed genes lacking enhancer regulation (Fig. 1B) despite having similar expression levels (Extended data Fig. 1F). To control for factors that might confound this analysis, we performed multivariate logistic regression using SE-status, Expression level, GC content, and gene length as predictors for transcription stress at genes. The results of this model demonstrate that SE-status was the most statistically significant independent predictor of transcription stress (Extended Data Fig. 1G), indicating specificity to genes regulated by SEs.

DSBs at these genes coincide with TOP1cc and R-loops (Fig. 1C). Analysis of sBLISS data for MCF7 cells depleted for TOP1 and R-loops shows that most of these genes have significantly fewer breaks upon TOP1 and R-loops depletion (Extended data Fig. 2A). Additionally, DSB are enriched at SE-regulated genes in both cycling and G1-synchronized cells (Extended data Fig. 2C-F), validating their transcriptional origin and independence from replication.

The relationship between transcription and DSBs was previously established[13–15]. However, our results suggest that this association may be more assertive in SE-regulated genes. This prompted us to further investigate the discrepancies in the relationship between transcription and DSBs in SE-regulated versus non-SE-regulated genes. Interestingly, the strong correlation between gene expression and DSBs is only observed in SE-regulated genes and not in others (Extended Data Fig. 2G-I). Moreover, highly expressed genes enriched with DSBs showed significantly more enrichment in TOP1, TOP1cc, and R-loops than similarly expressed genes lacking DSBs (Extended Data Fig. 3A-E). Additionally, SE-regulated genes were significantly more prevalent in highly expressed genes enriched with DSBs (Extended Data Fig. 3F). These findings indicate that the association between transcription and DSBs is primarily attributed to SE-regulated genes.

122 Since SE activity is significantly more potent in cancer cells than in normal cells, we next aimed
123 to test whether the enrichment of DSBs at SE-regulated genes is a phenomenon specific to
124 cancer. To investigate this, we mapped the Breakome in 10 different cell lines, which included
125 three normal cell lines, one pre-malignant cell line, and six cancer cell lines. These results
126 indicate that SE-regulated genes are more enriched with DSBs in cancer cells than in their
127 normal counterparts (Fig. 1D), demonstrating the specificity of transcriptional stress to active
128 SE-regulated oncogenes in cancer cells.
129
130 **The super-enhancer landscape shapes transcription stress sites in cancer.**
131 The landscape of SEs is cell-type specific, enhancing the transcription of different genes based
132 on cell identity[12]. If SEs cause transcription stress in their target genes, one would predict that
133 various cancer cell types exhibit distinct transcription stress sites reflective of their SE
134 landscapes. To test this, we mapped the Breakome of multiple cancer cell lines. For each cell
135 line, SE-regulated genes were identified, common SE-regulated genes from other cell lines were
136 excluded, and low to moderately expressed genes, as well as genes with low break counts, were
137 filtered out. Interestingly, SE-regulated genes were enriched in DSBs, predominantly in the cell
138 type where they are regulated by an SE (Fig. 2A). This pattern was also apparent when
139 comparing cell types with very similar characteristics (MCF7 and T47D, both of which are ER+
140 luminal A breast cancer cell types), yet exhibit differing SE landscapes, suggesting that
141 transcription stress sites correspond to the SE landscape in cancer.
142 If SE activity indeed causes the accumulation of DSBs at transcription stress sites, then inhibiting
143 their activity would abolish SE-mediated DSB accumulation, while conversely, activating it would
144 enhance DSB accumulation at target genes. To test this hypothesis, we employed two
145 strategies. First, we inhibited SE activity in MCF7 and HeLa cells through two approaches, using
146 the BRD4 degron dBET6 and the CDK9 inhibitor (CDK9i) NVP-2, both of which have been shown
147 to lead to SE collapse and repress SE-mediated transcription[32,33]. Examining the breakome after
148 a 4-hour incubation with these inhibitors, which effectively suppressed SE-regulated
149 transcriptional activity as measured by spike-in real-time quantitative PCR (RT-qPCR)
150 (Extended data Fig. 3G-H), revealed that SE inhibition significantly reduced DSBs at SE-
151 regulated genes in both cell lines (Fig. 2B). Furthermore, SE inhibition using either dBET6 or
152 CDK9i eliminated the cell-type-specific effect of the SE landscape on the breakome (Fig. 2C).

153 Secondly, we utilized the fact that MCF7 cells are estrogen receptor-positive and that many of
154 their SEs are estrogen-responsive and analyzed the breakome of these cells before and after
155 estrogen addition. We specifically focused on genes regulated by estrogen receptor-occupied
156 SE-regulated genes (ERSE-genes), which were previously characterized[34]. Consistently, ERSE
157 induction significantly increased DSBs in target genes but not random genes (Fig. 2 D-E),
158 indicating that SE activity promotes the accumulation of DSBs at SE-regulated genes. Overall,
159 these data demonstrate that the landscape of SE activity in cancer cells shapes the landscape
160 of transcription stress in a cell-type and cancer-type-specific manner.
161

162 **Specificity of transcription stress at super-enhancer-regulated genes**

163 While multiple studies have demonstrated the formation of DSBs as a consequence of
164 transcription in general[35–37], our results indicate that this phenomenon is markedly amplified
165 when transcription is driven by SEs (Fig. 1B; Extended Data Fig. 1E–F). To further support this
166 observation, we employed two complementary approaches.

167 First, we examined predictors for the genomic location of transcription stress sites. Using SE-
168 regulated genes, typical enhancer (TE)-regulated genes, and general highly expressed genes
169 as predictors, multivariate LASSO regression revealed that SE regulation was the strongest
170 predictor of transcription stress sites (odds ratio = 12.61; Fig. 3A), despite the highly expressed
171 gene group exhibiting the highest transcriptional levels overall (Fig. 3B). Notably, SE regulation
172 was also the best predictor for genomic regions enriched in DSBs (Fig. 3C).

173 Second, to experimentally assess whether transcriptional activation through SEs promotes
174 greater transcription stress and DSB accumulation, we performed a deeper analysis on MCF7
175 breakome before and after estradiol treatment. Estrogen activates transcription via both
176 promoter-proximal binding and estrogen-responsive SEs[38], enabling us to compare the impact
177 of these two mechanisms. We grouped genes by their level of transcriptional upregulation
178 (based on GRO-seq fold change) and further subdivided each group into SE-regulated and non-
179 SE-regulated categories. Consistently, SE-regulated genes accumulated significantly more
180 DSBs than their non-SE counterparts. This effect was especially prominent in the highest
181 induction group (fold change > 10; Fig. 3D). A similar pattern was observed upon CDK9
182 inhibition, where SE-regulated genes exhibited a significantly greater reduction in DSBs

183 compared to non-SE-regulated genes with matched transcriptional downregulation (measured
184 by RNA-seq) (Fig. 3E).

185 Together, these results demonstrate that SE-mediated transcription is a major determinant of
186 transcription stress site formation and leads to amplified DSB accumulation compared to non-
187 SE transcriptional activity, even at equivalent expression levels.[38]
188

189 **γH2AX marks transcription stress sites**

190 Since DSBs are transcription-blocking lesions[28], transcription stress sites must be rapidly and
191 efficiently repaired to maintain high expression levels. γH2AX is a marker for DDR activation and
192 is among the first responders to DSBs, as it is phosphorylated within a few minutes after DSB
193 formation[39]. To further explore the role of DDR activation in the signaling of DSB repair at
194 transcription stress sites, we conducted ChIP-seq for γH2AX on untreated MCF7 cells and
195 analyzed high-quality available γH2AX ChIP-seq data for HeLa and IMR90 cell lines.
196 In both MCF7 and HeLa cells, γH2AX signal is predominantly localized to gene bodies, whereas
197 in IMR90 cells the distribution is more balanced between genic and intergenic regions (Extended
198 Data Fig. 4A). This pattern is consistent with a greater burden of transcription-associated stress
199 in cancer cells compared to non-transformed cells. Moreover, γH2AX is distributed across the
200 gene bodies of top targets (Extended Data Fig. 4B-D). Ranking genes based on their γH2AX
201 enrichment revealed that most exhibit relatively low γH2AX levels, while a subset displays high
202 to very high γH2AX density (Fig. 4A). This finding indicates a non-uniform distribution of γH2AX,
203 with preferential enrichment at specific genes. Interestingly, all transcription stress sites are
204 among the top 8% of γH2AX targets (Fig. 4A-B). This enrichment in the top γH2AX genes was
205 much less for the general high DSB genes (Fig. 4C). Additionally, top γH2AX targets are enriched
206 in SE-regulated genes across each respective cell line (P= $2.3e^{-41}$, $2.9e^{-39}$, $8.5e^{-8}$, for MCF7,
207 HeLa, and IMR90 respectively; hypergeometric test) (Fig. 4D-F).
208 The observed γH2AX enrichment across the three cell lines was consistent for several
209 oncogenes, including miR21, NEAT1, MALAT1, and ACTG1 (Fig. 4G). Interestingly, many
210 prominent γH2AX targets were unique to specific cell lines, most with established activity in their
211 respective cell types (Fig. 4H). For example, two top γH2AX targets in MCF7, TFF1 and GATA3,
212 are well-known oncogenic transcription factors primarily active in breast cancer and were

213 exclusively enriched in γH2AX in MCF7 cells. KRT7 and CCAT1, which are overexpressed and
214 linked to poor prognosis in cervical cancer, are enriched in γH2AX only in HeLa cells.
215 Consequently, we questioned whether these variations in γH2AX enrichment stem from cell-
216 type-specific SEs. To investigate this, we compared γH2AX enrichment of top targets that are
217 regulated by a cell-type specific SE in each respective cell line. Indeed, SE-regulated γH2AX
218 targets are predominantly enriched in the cell line where they are regulated by an SE (Fig. 4 I),
219 underscoring that γH2AX marks cell type-specific transcriptional stress sites.
220
221 **γH2AX levels vary significantly across genes enriched with high endogenous DSBs**.
222 Next, we examined the correlation between DSBs and γH2AX in genes. Interestingly, analyzing
223 the correlation between each gene's breaks and its γH2AX density revealed high variability of
224 γH2AX among genes with high DSBs (Fig. 4 J-K and Extended Data Fig. 5A-E). Some genes
225 enriched with breaks are marked by γH2AX while others are not, in both MCF7 and HeLa cell
226 lines (Extended Data Fig. 5F-G). This discrepancy in γH2AX enrichment at various DSB-
227 enriched genes might suggest that regions enriched with endogenous DSBs in the genome are
228 signaled with varying efficiency and potentially repaired unevenly, depending on their context.
229
230 **γH2AX preferentially marks endogenous DSBs of active chromatin**
231 Although genes exhibit a moderate correlation between DSB density and γH2AX density
232 (Spearman's correlation = 0.57), non-linear regression analysis using Generalized Additive
233 Models (GAM) indicates that the transcription stress score explains a greater proportion of the
234 variance in γH2AX levels (R^2 = 0.73) than any individual component (Fig. 5A and Extended data
235 Fig. 6). To further investigate the underlying factors contributing to variability in γH2AX at high
236 DSBs genes, we defined two groups of genes with similar break density but different γH2AX
237 levels (Fig. 4J): $DSBs^{high}$ $\gamma H2AX^{high}$ genes and $DSBs^{high}$ $\gamma H2AX^{low}$ genes. $DSBs^{high}$ $\gamma H2AX^{high}$
238 genes are more expressed and enriched with transcription stress markers (TOP1, TOP1cc, and
239 R-loops) (Extended data Fig. 7A-E). Furthermore, 64 of the 65 transcriptional stress sites were
240 among $DSBs^{high}$ $\gamma H2AX^{high}$ genes (Fig. 4J and Fig. 5B). This group also included all SE-regulated
241 transcription stress sites and is significantly enriched with active SE-regulated genes compared

242 to DSBs^{high} γH2AX^{low} genes, which lack active SE-regulated genes and transcriptional stress
243 hotspots (Fig. 5C and Extended data Fig. 7F).
244 Chromatin context was previously shown to impact the phosphorylation of H2AX, specifically,
245 heterochromatin was shown to be refractory to H2AX phosphorylation[40]. Therefore, we
246 wondered whether the observed DSBs^{high} γH2AX^{low} genes are close to, or within,
247 heterochromatic regions. To gain further insights into whether γH2AX preferentially marks
248 endogenous DSBs based on the chromatin context, we defined two groups of individual DSB
249 sites across the genome according to their γH2AX signal (Fig. 5D and Extended data Fig. 7G)
250 and mapped them to the different chromatin states of MCF7 and HeLa. γH2AX^{high} DSBs were
251 biased toward active chromatin states, while γH2AX^{low} DSBs were biased toward silent regions
252 (Fig. 5E and Extended data Fig. 7H). These results made us question whether the differential
253 γH2AX marking at DSB-enriched genes was due to their chromatin context. Indeed, the DSBs^{high}
254 γH2AX^{low} group, but not the DSBs^{high} γH2AX^{high} group, is enriched with DSB-enriched genes that
255 overlap with heterochromatin and repressed chromatin states (Fig. 5F and Extended data Fig.
256 7I). Additionally, an analysis of the chromatin state contexts of DSBs^{high} γH2AX^{high} and DSBs^{high}
257 γH2AX^{low} in both MCF7 and HeLa showed that DSBs^{high} γH2AX^{high} genes primarily overlap with
258 chromatin regions associated with strong transcriptional activity. Conversely, DSBs^{high} γH2AX^{low}
259 has a higher overlap with heterochromatin, repressed, and weak transcription chromatin states
260 (Fig. 5G and Extended data Fig. 7J).
261 Our results illustrate that DSBs alone do not fully account for the variation in γH2AX across genes
262 (Fig. 5A). Conducting GAM on genes categorized by their chromatin context revealed that DSBs
263 primarily account for variation in γH2AX in genes that overlap with strong enhancer regions and
264 least in genes that overlap with repressed or heterochromatin regions (Fig. 5H). This suggests
265 that breaks and γH2AX are most closely associated with genes spanning strong enhancer
266 elements.
267 Collectively, these results indicate that γH2AX preferentially marks endogenous DSBs in active
268 chromatin compared to inactive chromatin. This differential marking may influence variations in
269 the repair efficiency of distinct DSBs based on their chromatin context and γH2AX signalling.
270
271 **Genes marked with γH2AX exhibit high DSB turnover and repair efficiency**

272 sBLISS, like other methods used to map DSBs, captures a snapshot of the DSB landscape at
273 the moment of fixation[41–45]. Therefore, a gene that shows high levels of DSBs could experience
274 frequent DSB generation and repair cycles (high DSB turnover) or infrequent DSB generation
275 and repair cycles (low DSB turnover). To estimate the rate of endogenous DSB repair across
276 the genome, we mapped DSBs in HeLa cells after disrupting the two major DSB repair pathways,
277 non-homologous end-joining (NHEJ)[45] and homologous recombination (HR), by using siRNAs
278 that target XRCC4 and RAD51, respectively (Extended Data Fig. 8A). We applied DESeq2 [46] to
279 the DSB count matrices to identify genes with significant DSB accumulation following knockdown
280 of each factor. Hierarchical clustering of these genes revealed three distinct response groups:
281 Cluster 1 genes were primarily affected by XRCC4 depletion, Cluster 2 by both XRCC4 and
282 RAD51 (with a stronger effect from RAD51), and Cluster 3 mainly by RAD51 depletion (Extended
283 Data Fig. 8B). Importantly, no cluster was exclusively dependent on RAD51, consistent with
284 previous reports that XRCC4 localizes to both RAD51-bound and unbound breaks[47].
285 Further analysis revealed that all three clusters are associated with transcription stress, with
286 Cluster 1 (XRCC4-dependent genes) showing the strongest enrichment (Extended Data Fig.
287 8C). These genes also exhibit the highest baseline break density (Extended Data Fig. 8D),
288 suggesting that NHEJ is particularly engaged at the most DSB-prone, transcriptionally stressed
289 loci. Collectively, these findings support the involvement of both NHEJ and HR in repairing
290 transcription-associated DSBs, while indicating a prominent role for XRCC4-mediated repair at
291 the most vulnerable genomic sites. Based on this, we proceeded to quantify genome-wide DSB
292 turnover using XRCC4 knockdown‑induced break accumulation as a proxy for repair efficiency.
293 Regions that are frequently repaired should accumulate excessive DSBs due to repair
294 impairment. In contrast, regions with inherently low repair rates should show minimal DSB
295 accumulation (Fig. 6A). 429 genes showed significant DSB accumulation upon XRCC4 KD (Fig.
296 6B). Stratifying these genes by their baseline DSB density revealed a trend in which groups with
297 higher baseline DSB density showed progressively greater impact following XRCC4 KD (Fig.
298 6C), supporting the notion that DSB-prone genes are repaired more effectively. Enrichment
299 analysis indicated a strong enrichment of DSBs^{high} γH2AX^{high} genes (Fig. 6D), along with the top
300 DSB-enriched and active SE-regulated genes. In contrast, DSBs^{high} γH2AX^{low} genes and the
301 bottom DSB-enriched and non-SE-regulated genes were not significantly enriched. Directly
302 examining the effects of repair impairment demonstrated that NHEJ impairment leads to the

303 most pronounced accumulation of DSBs at DSBs^{high} γH2AX^{high} and SE-regulated genes,
304 compared to DSBs^{high} γH2AX^{low} and non-SE genes (Fig. 6EE). To investigate further how γH2AX
305 patterns influence the repair turnover of endogenous DSBs, we used the accumulation of DSBs
306 following repair impairment as a proxy for repair efficiency (Fig. 6F). Within DSBs^{high} γH2AX^{high}
307 genes, repair efficiency markedly increases with higher DSB levels; however, within DSBs^{high}
308 γH2AX^{low} genes, this relationship of repair efficiency is significantly weaker (Fig. 6G). These
309 findings suggest two distinct types of endogenous DSBs: γH2AX-marked DSBs, which exhibit
310 high turnover, and poorly γH2AX-marked DSBs, which exhibit low turnover.
311
312 **Efficient repair at transcriptional stress sites is crucial for SE-regulated oncogene**
313 **expression.**
314 High γH2AX marking and repair efficiency at transcription stress sites under physiological
315 conditions suggest that these regions rely on DSB repair to sustain their elevated transcriptional
316 levels. To test this, we inhibited DNA-PKcs, a core component of the canonical NHEJ pathway,
317 using a selective small-molecule inhibitor (DNA-PKci) in MCF7 cells. We then performed sBLISS
318 to map genome-wide DSBs and spike-in RT-qPCR to quantify transcript levels.
319 To assess genome-wide DSB accumulation, we divided the genome into 10 kb bins, quantified
320 DSBs per bin, and used DESeq2 to identify bins with significant DSB enrichment in DNA-PKci-
321 treated cells. We then overlapped significantly differential bins with annotated gene regions to
322 determine whether specific gene classes were disproportionately affected. Indeed, significantly
323 upregulated bins were strongly enriched at SE-regulated genes and not at non-SE-regulated
324 genes (Extended Data Fig. 9A–b).
325 Measuring mRNA levels of the top γH2AX marked SE-regulated genes upon DNA-PKci indicated
326 a significant decrease in their expression levels (Extended data Fig. 9C). To determine whether
327 the observed transcriptional reduction was due to repair failure rather than secondary signaling
328 effects, we also knocked down XRCC4 under the same conditions as in Fig. 6. This similarly led
329 to downregulation of highly γH2AX-marked SE-regulated genes that accumulated DSBs
330 (Extended Data Fig. 9D). Together, these results demonstrate that efficient DSB repair via the
331 NHEJ pathway is required to sustain the transcription of SE-regulated oncogenes.
332

333 **Efficiently repaired transcriptional DSBs cluster in the 3D genome.**

334 Induced DSBs cluster in repair foci[48,49], especially when introduced within active genes[49]. We
335 hypothesized that endogenous transcription-associated DSBs might also cluster within the 3D
336 genome as a mechanism to enhance repair. To investigate this, we analyzed high-quality Hi-C
337 data generated from MCF7 cells[50].

338 To quantify the clustering of genes associated with break density, we introduced two Hi-C
339 metrics that measure how a group of genes with a common characteristic (e.g., break density)
340 is clustered (Fig. 7A): intra-group interactions, which represent the sum of genomic interaction
341 scores between genes within the same group, and inter-group interactions, the sum of genomic
342 interaction scores between genes from different groups. A clustered group of genes is expected
343 to exhibit higher intra-group interactions and lower inter-group interactions compared to a non-
344 clustered group (Fig. 7A). To investigate whether DSB-enriched genes exhibit clustering,
345 genomic bins were arranged according to the break densities of their constituent genes and
346 categorized into nine groups based on break density. Indeed, bins containing DSB-enriched
347 genes demonstrate higher intra-group interactions and lower inter-group interactions than bins
348 with low DSB genes (Fig. 7B), indicating the clustering of high DSB genes.

349 The percentage of strong interactions among $DSBs^{high}$ $\gamma H2AX^{high}$ genes is significantly higher
350 than that among $DSBs^{high}$ $\gamma H2AX^{low}$ genes (Fig. 7C and Extended data Fig. 9A). Additionally,
351 transcription stress sites exhibit a substantially higher percentage of strong interactions
352 compared to random sites with similar chromosomal distribution (Fig. 7 D and Extended data
353 Fig. 9B), suggesting that efficiently repaired transcription stress sites are clustered together. To
354 avoid the confounding effects of SE regulation on 3D interactions, we analysed Hi-C data from
355 the MCF10A cell line, which was conducted parallel to MCF7[50]. MCF10A is a pre-cancerous cell
356 line that shares many common SE-regulated genes with MCF7 cells. However, these common
357 genes are significantly more enriched in DSBs within MCF7 cells (Fig. 7E). This enabled a
358 comparison of the same regions with the same SE-regulation status but differing DSB
359 enrichment. Indeed, bins containing common SE-regulated genes, as opposed to random
360 regions, showed significantly higher strong interchromosomal interactions in MCF7 cells than in
361 MCF10A cells (Fig. 7F-G), indicating that transcription stress sites undergo increased clustering.
362

363 **The repair of transcription stress sites is mutagenic.**

364 Since transcription stress sites are more likely to be in close proximity (Fig. 7), we suspected
365 that recurrent DSBs at these sites could serve as mutational hotspots. To explore this, we
366 downloaded and analysed mutational data from the COSMIC database[51], which contains
367 extensive mutational data on human cancers. While mutations within coding regions are subject
368 to selective pressures based on their impact on protein function and cellular fitness [52,53], non-
369 coding mutations—particularly those in intronic regions—are less constrained by selection and
370 can provide a more direct readout of a region's intrinsic susceptibility to mutation. To minimize
371 the confounding effects of selection, we focused on non-coding intronic mutations in breast
372 cancer and intersected them with our breakome data from the breast cancer cell line MCF7.
373 Grouping genes by their DSB or γH2AX density revealed that genes with a higher break or
374 γH2AX density exhibit greater intronic mutation density (Fig. 8A-C). This association suggests
375 that the repair of transcription-associated DSBs is inherently error-prone, leading to the
376 accumulation of mutations over time.
377
378 To disentangle the effects of DSB density from repair and turnover, we modelled the relationship
379 between γH2AX density and break density using nonlinear least squares (NLS) regression. The
380 residuals from this model were then used to classify genes in each group into three subgroups:
381 Under-responsive genes (negative residuals), moderately responsive genes (residuals near
382 zero), and over-responsive genes (positive residuals). Notably, over-responsive genes showed
383 the most increase in mutation density with increased DSBs (Fig. 8D-F). This suggests that
384 excessive γH2AX accumulation at DSB sites, which reflects recurrent DSB formation and
385 turnover, and heightened repair activity, is error-prone. Furthermore, $DSBs^{high}$ $\gamma H2AX^{high}$ genes,
386 but not $DSBs^{high}$ $\gamma H2AX^{low}$ genes, demonstrated intronic mutation enrichment, compared to
387 randomly selected genes (Fig. 8G-I). Collectively, these results suggest that transcription stress
388 sites are prone to mutations, likely due to a combination of high DSB turnover, clustering, and
389 defective repair. These findings have important implications for understanding how transcription-
390 associated genome instability contributes to cancer evolution, as recurrent DSB formation and
391 error-prone repair may drive the accumulation of oncogenic mutations over time.
392
393
394

395 **Discussion**

396 Cellular processes–particularly those upregulated in cancer such as replication and
397 transcription–play a significant role in genomic instability[54]. Replication-dependent instability,
398 caused by replication stress[4,55], has been extensively studied, revealing its landscape[56,57],
399 causes, consequences[58], and potential for therapeutic targeting[59,60]. Transcription-replication-
400 dependent genomic instability[61], particularly transcription-replication conflicts (TRCs)[62], has also
401 been well characterized, with studies examining its genomic distribution, influencing factors, and
402 the impact of transcription-replication orientation[63]. Recently, attention has shifted toward
403 transcription-dependent, replication-independent genomic instability, namely, transcriptional
404 DSBs. While research has begun to uncover their causes and implications in disease[17,64],
405 several key aspects remain underexplored, including genomic hotspot patterns, repair
406 mechanisms, and their impact on cancer evolution. By utilizing multi-omic approaches, we
407 delineated the landscape of transcription stress and revealed that SEs shape the transcriptional
408 program and contribute to the formation of genomic instability hotspots in cancer.

409 SEs determine cell identity in normal cells[31]. In cancer cells, SEs enhance the transcription of
410 oncogenes to promote cell growth and proliferation [6,8,65,66]. Previous studies have established
411 that highly transcribed genes are prone to transcription-associated genome instability[13,15,20,30,67].
412 However, our data reveal that this relationship is significantly amplified in SE-regulated genes,
413 suggesting that transcriptional stress is not merely a function of high gene expression but is
414 modulated by SE activity. The enrichment of highly expressed SE-regulated genes among
415 transcription stress sites was over 30-fold higher than expected by chance, far exceeding that
416 of highly expressed TE-regulated genes or highly expressed genes lacking enhancer regulation.
417 This specificity suggests that SEs impose a unique transcriptional burden that predisposes their
418 target genes to DNA damage. The functional implications are significant, as many SE-regulated
419 genes encode critical oncogenes, including CCND1, NEAT1, miR21, and MYC, which play
420 essential roles in tumorigenesis.

421 The pronounced enrichment of DSBs at SE-regulated oncogenes was considerably higher in
422 cancer cell lines compared to their normal counterparts, indicating that SEs primarily drive
423 genomic instability in cancer cells. SEs have previously been recognized as hyper-mutated
424 regions that can alter the expression of target genes and contribute to carcinogenesis[6]. Our

425 findings advance this understanding by demonstrating that SEs also dictate mutational
426 susceptibility at their target genes by causing hyper-transcription.
427 Although sBLISS captures a snapshot of the DSB landscape at the time of fixation, we could
428 estimate the repair and turnover of DSBs across genes, revealing how transcription stress sites
429 particularly undergo high DSB turnover and repair efficiency. To accommodate this, these
430 regions experience constant activation of the DNA damage response (DDR) and H2AX
431 phosphorylation. Interestingly, not all DSB-enriched genes exhibit such efficient repair signaling,
432 suggesting two distinct types of DSB-bearing genes in the genome: genes that undergo frequent
433 breakage-repair cycles and are actively recognized by the repair machinery, and genes that,
434 despite being DSB-rich, experience infrequent breakage-repair cycles and remain persistent.
435 These persistent DSBs, found in regions with low expression, have a lower demand for repair.
436 We speculate that these latter genes experience breakage in a more programmed manner
437 compared to genes with high turnover, which break as a byproduct of transcription. The nature
438 of these DSBs requires further investigation.
439 Additionally, we demonstrate that genes with efficient DSB repair signaling are more vulnerable
440 to mutations than genes with comparable DSB levels but less effective repair signaling,
441 emphasizing the importance of the repair process itself in mutagenesis at transcription stress
442 sites.
443 In summary, our findings highlight SEs as critical determinants of transcriptional stress and
444 genomic instability in cancer. By integrating multi-omic data, we demonstrate that SE-driven
445 hyper-transcription uniquely predisposes oncogenes to DNA breakage and mutation, revealing
446 a previously underappreciated mechanism by which SEs shape the cancer genome. These
447 insights not only deepen our understanding of transcription-dependent DNA damage but also
448 point to SE-associated transcription stress as a potential vulnerability that could be
449 therapeutically exploited. Further investigation into the distinct dynamics of DSB repair at these
450 loci may uncover new strategies to modulate genome stability in cancer.
451

References

1. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* vol. 144 646–674 Preprint at https://doi.org/10.1016/j.cell.2011.02.013 (2011).
2. Bradner, J. E., Hnisz, D. & Young, R. A. Transcriptional Addiction in Cancer. *Cell* vol. 168 629–643 Preprint at https://doi.org/10.1016/j.cell.2016.12.013 (2017).
3. Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* vol. 152 1237–1251 Preprint at https://doi.org/10.1016/j.cell.2013.02.014 (2013).
4. Bowry, A., Kelly, R. D. W. & Petermann, E. Hypertranscription and replication stress in cancer. *Trends in Cancer* vol. 7 863–877 Preprint at https://doi.org/10.1016/j.trecan.2021.04.006 (2021).
5. Pott, S. & Lieb, J. D. What are super-enhancers? *Nature Genetics* vol. 47 8–12 Preprint at https://doi.org/10.1038/ng.3167 (2015).
6. Bal, E. *et al.* Super-enhancer hypermutation alters oncogene expression in B cell lymphoma. *Nature* **607**, 808–815 (2022).
7. Wang, X., Cairns, M. J. & Yan, J. Super-enhancers in transcriptional regulation and genome organization. *Nucleic Acids Research* vol. 47 11481–11496 Preprint at https://doi.org/10.1093/nar/gkz1038 (2019).
8. Wang, M. T. *et al.* Super-enhancers complexes zoom in transcription in cancer. *Journal of Experimental and Clinical Cancer Research* vol. 42 Preprint at https://doi.org/10.1186/s13046-023-02763-5 (2023).
9. Bacabac, M. & Xu, W. Oncogenic super-enhancers in cancer: mechanisms and therapeutic targets. *Cancer and Metastasis Reviews* vol. 42 471–480 Preprint at https://doi.org/10.1007/s10555-023-10103-4 (2023).
10. Hnisz, D. *et al.* Convergence of Developmental and Oncogenic Signaling Pathways at Transcriptional Super-Enhancers. *Mol Cell* **58**, 362–370 (2015).
11. Morton, A. R. *et al.* Functional Enhancers Shape Extrachromosomal Oncogene Amplifications. *Cell* **179**, 1330-1341.e13 (2019).
12. Hnisz, D. *et al.* XSuper-enhancers in the control of cell identity and disease. *Cell* **155**, 934 (2013).
13. Hidmi, O., Oster, S., Monin, J. & Aqeilan, R. I. TOP1 and R-loops facilitate transcriptional DSBs at hypertranscribed cancer driver genes. *iScience* **27**, 109082 (2024).
14. Lensing, S. V. *et al.* DSBCapture: In situ capture and sequencing of DNA breaks. *Nat Methods* **13**, 855–857 (2016).
15. Oster, S. & Aqeilan, R. I. Programmed DNA Damage and Physiological DSBs: Mapping, Biological Significance and Perturbations in Disease States. *Cells* vol. 9 Preprint at https://doi.org/10.3390/cells9081870 (2020).
16. Girasol, M. J. *et al.* RAD51-mediated R-loop formation acts to repair transcription-associated DNA breaks driving antigenic variation in Trypanosoma brucei. *Proc Natl Acad Sci U S A* **120**, (2023).
17. Stork, C. T. *et al.* Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. (2016) doi:10.7554/eLife.17548.001.
18. Sollier, J. *et al.* Transcription-Coupled Nucleotide Excision Repair Factors Promote R-Loop-Induced Genome Instability. *Mol Cell* **56**, 777–785 (2014).
19. Halazonetis, T. D., Gorgoulis, V. G. & Bartek, J. *An Oncogene-Induced DNA Damage Model for Cancer Development*. https://www.science.org.

20. Cristini, A. *et al.* Dual Processing of R-Loops and Topoisomerase I Induces Transcription-Dependent DNA Double-Strand Breaks. *Cell Rep* **28**, 3167-3181.e6 (2019).
21. Milano, L., Gautam, A. & Caldecott, K. W. DNA damage and transcription stress. *Molecular Cell* vol. 84 70–79 Preprint at https://doi.org/10.1016/j.molcel.2023.11.014 (2024).
22. Min, J. *et al.* Mechanisms of insertions at a DNA double-strand break. *Mol Cell* **83**, 2434-2448.e7 (2023).
23. Reijns, M. A. M. *et al.* Signatures of TOP1 transcription-associated mutagenesis in cancer and germline. *Nature* **602**, 623–631 (2022).
24. Nakazawa, Y. *et al.* Ubiquitination of DNA Damage-Stalled RNAPII Promotes Transcription-Coupled Repair. *Cell* **180**, 1228-1244.e24 (2020).
25. Gyenis, A. *et al.* Genome-wide RNA polymerase stalling shapes the transcriptome during aging. *Nat Genet* **55**, 268–279 (2023).
26. Saponaro, M. *et al.* RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. *Cell* **157**, 1037–1049 (2014).
27. Min, S., Ji, J. H., Heo, Y. & Cho, H. Transcriptional regulation and chromatin dynamics at DNA double-strand breaks. *Experimental and Molecular Medicine* vol. 54 1705–1712 Preprint at https://doi.org/10.1038/s12276-022-00862-5 (2022).
28. Machour, F. E. & Ayoub, N. Transcriptional Regulation at DSBs: Mechanisms and Consequences. *Trends in Genetics* vol. 36 981–997 Preprint at https://doi.org/10.1016/j.tig.2020.01.001 (2020).
29. Crossley, M. P., Bocek, M. & Cimprich, K. A. R-Loops as Cellular Regulators and Genomic Threats. *Molecular Cell* vol. 73 398–411 Preprint at https://doi.org/10.1016/j.molcel.2019.01.024 (2019).
30. Cristini, A. *et al.* DNA-PK triggers histone ubiquitination and signaling in response to DNA double-strand breaks produced during the repair of transcription-blocking topoisomerase I lesions. *Nucleic Acids Res* **44**, 1161–1178 (2016).
31. Wang, Y. *et al.* SEdb 2.0: a comprehensive super-enhancer database of human and mouse. *Nucleic Acids Res* **51**, D280–D290 (2023).
32. Shu, S. *et al.* Synthetic Lethal and Resistance Interactions with BET Bromodomain Inhibitors in Triple-Negative Breast Cancer. *Mol Cell* **78**, 1096-1113.e8 (2020).
33. Winter, G. E. *et al.* BET Bromodomain Proteins Function as Master Transcription Elongation Factors Independent of CDK9 Recruitment. *Mol Cell* **67**, 5-18.e19 (2017).
34. Zheng, Z. Z. *et al.* Super-enhancer-controlled positive feedback loop BRD4/ERα-RET-ERα promotes ERα-positive breast cancer. *Nucleic Acids Res* **50**, 10230–10248 (2022).
35. Geraud, M. *et al.* TDP1 mutation causing SCAN1 neurodegenerative syndrome hampers the repair of transcriptional DNA double-strand breaks. *Cell Rep* **43**, (2024).
36. Bunch, H. *et al.* Transcriptional elongation requires DNA break-induced signalling. *Nat Commun* **6**, (2015).
37. Madabhushi, R. *et al.* Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. *Cell* **161**, 1592–1605 (2015).
38. Carroll, J. S. *et al.* Genome-wide analysis of estrogen receptor binding sites. *Nat Genet* **38**, 1289–1297 (2006).
39. γH2AX and cancer.

542 40. Kim, J. A., Kruhlak, M., Dotiwala, F., Nussenzweig, A. & Haber, J. E. Heterochromatin is
543 refractory to γ-H2AX modification in yeast and mammals. *Journal of Cell Biology* **178**,
544 209–218 (2007).
545 41. Bouwman, B. A. M. *et al.* Genome-wide detection of DNA double-strand breaks by in-
546 suspension BLISS. *Nat Protoc* **15**, 3894–3941 (2020).
547 42. Yan, W. X. *et al.* BLISS is a versatile and quantitative method for genome-wide profiling
548 of DNA double-strand breaks. *Nat Commun* **8**, (2017).
549 43. Canela, A. *et al.* DNA Breaks and End Resection Measured Genome-wide by End
550 Sequencing. *Mol Cell* **63**, 898–911 (2016).
551 44. Dobbs, F. M. *et al.* Precision digital mapping of endogenous and induced genomic DNA
552 breaks by INDUCE-seq. *Nat Commun* **13**, (2022).
553 45. Wei, P. C. *et al.* Long Neural Genes Harbor Recurrent DNA Break Clusters in Neural
554 Stem/Progenitor Cells. *Cell* **164**, 644–655 (2016).
555 46. Anders, S. & Huber, W. Differential expression analysis for sequence count data.
556 *Genome Biol* **11**, (2010).
557 47. Aymard, F. *et al.* Transcriptionally active chromatin recruits homologous recombination
558 at DNA double-strand breaks. *Nat Struct Mol Biol* **21**, 366–374 (2014).
559 48. Arnould, C. *et al.* Chromatin compartmentalization regulates the response to DNA
560 damage. *Nature* **623**, 183–192 (2023).
561 49. Aymard, F. *et al.* Genome-wide mapping of long-range contacts unveils clustering of
562 DNA double-strand breaks at damaged active genes. *Nat Struct Mol Biol* **24**, 353–361
563 (2017).
564 50. Barutcu, A. R. *et al.* Chromatin interaction analysis reveals changes in small
565 chromosome and telomere clustering between epithelial and breast cancer cells.
566 *Genome Biol* **16**, (2015).
567 51. Sondka, Z. *et al.* COSMIC: a curated database of somatic variants and clinical data for
568 cancer. *Nucleic Acids Res* **52**, D1210–D1217 (2024).
569 52. Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences
570 of genetic heterogeneity in cancer evolution. *Nature* vol. 501 338–345 Preprint at
571 https://doi.org/10.1038/nature12625 (2013).
572 53. Yates, L. R. & Campbell, P. J. Evolution of the cancer genome. *Nature Reviews*
573 *Genetics* vol. 13 795–806 Preprint at https://doi.org/10.1038/nrg3317 (2012).
574 54. Negrini, S., Gorgoulis, V. G. & Halazonetis, T. D. Genomic instability an evolving
575 hallmark of cancer. *Nature Reviews Molecular Cell Biology* vol. 11 220–228 Preprint at
576 https://doi.org/10.1038/nrm2858 (2010).
577 55. Hills, S. A. & Diffley, J. F. X. DNA replication and oncogene-induced replicative stress.
578 *Current Biology* vol. 24 Preprint at https://doi.org/10.1016/j.cub.2014.04.012 (2014).
579 56. Kim, J. *et al.* Replication Stress Shapes a Protective Chromatin Environment across
580 Fragile Genomic Regions. *Mol Cell* **69**, 36-47.e7 (2018).
581 57. Gaillard, H., García-Muse, T. & Aguilera, A. Replication stress and cancer. *Nature*
582 *Reviews Cancer* vol. 15 276–280 Preprint at https://doi.org/10.1038/nrc3916 (2015).
583 58. Zeman, M. K. & Cimprich, K. A. *R E V I E W Causes and Consequences of Replication*
584 *Stress*. (2014).
585 59. Machour, F. E. *et al.* Harnessing DNA replication stress to target RBM10 deficiency in
586 lung adenocarcinoma. *Nat Commun* **15**, (2024).

587 60. Cybulla, E. & Vindigni, A. Leveraging the replication stress response to optimize cancer
588 therapy. *Nature Reviews Cancer* vol. 23 6–24 Preprint at https://doi.org/10.1038/s41568-
589 022-00518-6 (2023).
590 61. Corazzi, L. *et al.* Linear interaction between replication and transcription shapes DNA
591 break dynamics at recurrent DNA break Clusters. *Nat Commun* **15**, (2024).
592 62. Hamperl, S. & Cimprich, K. A. Conflict Resolution in the Genome: How Transcription and
593 Replication Make It Work. *Cell* vol. 167 1455–1467 Preprint at
594 https://doi.org/10.1016/j.cell.2016.09.053 (2016).
595 63. Hamperl, S., Bocek, M. J., Saldivar, J. C., Swigut, T. & Cimprich, K. A. Transcription-
596 Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA
597 Damage Responses. *Cell* **170**, 774-786.e19 (2017).
598 64. Geraud, M. *et al.* TDP1 mutation causing SCAN1 neurodegenerative syndrome hampers
599 the repair of transcriptional DNA double-strand breaks. *Cell Rep* **43**, (2024).
600 65. Zhou, R. W. & Parsons, R. E. Etiology of super-enhancer reprogramming and activation
601 in cancer. *Epigenetics and Chromatin* vol. 16 Preprint at https://doi.org/10.1186/s13072-
602 023-00502-w (2023).
603 66. Huang, H. *et al.* Defining super-enhancer landscape in triple-negative breast cancer by
604 multiomic profiling. *Nat Commun* **12**, (2021).
605 67. Mosler, T. *et al.* R-loop proximity proteomics identifies a role of DDX41 in transcription-
606 associated genomic instability. *Nat Commun* **12**, (2021).
607 68. Suehnholz, S. P. *et al.* Quantifying the Expanding Landscape of Clinical Actionability for
608 Patients with Cancer. *Cancer Discov* **14**, 49–65 (2024).
609 69. Hidmi, O., Oster, S., Shatleh, D., Monin, J. & Aqeilan, R. I. Protocol for mapping
610 physiological DSBs using in-suspension break labeling in situ and sequencing. *STAR*
611 *Protoc* **5**, (2024).
612 70. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome.
613 *Nature* **489**, 57–74 (2012).
614 71. Luo, Y. *et al.* New developments on the Encyclopedia of DNA Elements (ENCODE) data
615 portal. *Nucleic Acids Res* **48**, D882–D889 (2020).
616 72. Greer, Y. E. *et al.* Mitochondrial Matrix Protease ClpP Agonists Inhibit Cancer Stem Cell
617 Function in Breast Cancer Cells by Disrupting Mitochondrial Homeostasis. *Cancer*
618 *Research Communications* **2**, 1144–1161 (2022).
619 73. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human
620 epigenomes. *Nature* **518**, 317–329 (2015).
621 74. Guo, C., Liu, S., Wang, J., Sun, M. Z. & Greenaway, F. T. ACTB in cancer. *Clinica*
622 *Chimica Acta* vol. 417 39–44 Preprint at https://doi.org/10.1016/j.cca.2012.12.012
623 (2013).
624 75. Wang, J. *et al.* HES1 promotes aerobic glycolysis and cancer progression of colorectal
625 cancer via IGF2BP2-mediated GLUT1 m6A modification. *Cell Death Discov* **9**, (2023).
626 76. Bermúdez, G. *et al.* Long Noncoding RNA TALAM1 Is a Transcriptional Target of the
627 RUNX2 Transcription Factor in Lung Adenocarcinoma. *Curr Issues Mol Biol* **45**, 7075–
628 7086 (2023).
629 77. Zhu, S. *et al.* MicroRNA-21 targets tumor suppressor genes in invasion and metastasis.
630 *Cell Res* **18**, 350–359 (2008).

78. Xie, S. J. *et al.* mascRNA and its parent lncRNA MALAT1 promote proliferation and metastasis of hepatocellular carcinoma cells by activating ERK/MAPK signaling pathway. *Cell Death Discov* **7**, (2021).
79. Ren, F. J., Cai, X. Y., Yao, Y. & Fang, G. Y. JunB: a paradigm for Jun family in immune response and cancer. *Frontiers in Cellular and Infection Microbiology* vol. 13 Preprint at https://doi.org/10.3389/fcimb.2023.1222265 (2023).
80. Li, Z. X. *et al.* MALAT1: A potential biomarker in cancer. *Cancer Manag Res* **10**, 6757–6768 (2018).
81. Ordway, J. M., Fenster, S. D., Ruan, H. & Curran, T. A transcriptome map of cellular transformation by the fos oncogene. *Mol Cancer* **4**, (2005).
82. Tang, Y., Peng, X., Huang, X. & Li, J. Actin gamma 1 is a critical regulator of pancreatic ductal adenocarcinoma. *Saudi Journal of Gastroenterology* **28**, 239–246 (2022).
83. Yu, X., Li, Z., Zheng, H., Chan, M. T. V. & Wu, W. K. K. NEAT1: A novel cancer-related long non-coding RNA. *Cell Proliferation* vol. 50 Preprint at https://doi.org/10.1111/cpr.12329 (2017).
84. Kyjacova, L. *et al.* IER2-induced senescence drives melanoma invasion through osteopontin. *Oncogene* **40**, 6494–6512 (2021).
85. Qie, S. & Sang, N. Stanniocalcin 2 (STC2): a universal tumour biomarker and a potential therapeutical target. *Journal of Experimental and Clinical Cancer Research* vol. 41 Preprint at https://doi.org/10.1186/s13046-022-02370-w (2022).
86. Xu, B. *et al.* LncRNA SNHG3, a potential oncogene in human cancers. *Cancer Cell International* vol. 20 Preprint at https://doi.org/10.1186/s12935-020-01608-x (2020).
87. Wang, S. *et al.* Chromatin licensing and DNA replication factor 1 as a potential prognostic and diagnostic biomarker for gastric cancer. *Sci Rep* **14**, (2024).
88. Zhao, X. *et al.* TEP SNORD12B, SNORA63, and SNORD14E as novel biomarkers for hepatitis B virus-related hepatocellular carcinoma (HBV-related HCC). *Cancer Cell Int* **24**, (2024).
89. Liang, Y. *et al.* HSPB1 facilitates chemoresistance through inhibiting ferroptotic cancer cell death and regulating NF-κB signaling pathway in breast cancer. *Cell Death Dis* **14**, (2023).
90. Perry, J. K., Kannan, N., Grandison, P. M., Mitchell, M. D. & Lobie, P. E. Are trefoil factors oncogenic? *Trends in Endocrinology and Metabolism* vol. 19 74–81 Preprint at https://doi.org/10.1016/j.tem.2007.10.003 (2008).
91. Asano, Y. *et al.* IER5 generates a novel hypo-phosphorylated active form of HSF1 and contributes to tumorigenesis. *Sci Rep* **6**, (2016).
92. El-Mezayen, H. *et al.* MicroRNA-25 Exerts an Oncogenic Function by Regulating the Ubiquitin Ligase Fbxw7 in Hepatocellular Carcinoma. *Ann Surg Oncol* **28**,.
93. Zhu, Y. *et al.* Identification of prothymosin alpha (PTMA) as a biomarker for esophageal squamous cell carcinoma (ESCC) by label-free quantitative proteomics and Quantitative Dot Blot (QDB). *Clin Proteomics* **16**, (2019).
94. Wang, C. *et al.* The Oncogenic Role of TNFRSF12A in Colorectal Cancer and Pan-Cancer Bioinformatics Analysis. *Cancer Res Treat* **57**, 212–228 (2025).
95. Wortel, I. M. N., van der Meer, L. T., Kilberg, M. S. & van Leeuwen, F. N. Surviving Stress: Modulation of ATF4-Mediated Stress Responses in Normal and Malignant Cells. *Trends in Endocrinology and Metabolism* vol. 28 794–806 Preprint at https://doi.org/10.1016/j.tem.2017.07.003 (2017).

96. Oji, Y. *et al.* The translation elongation factor eEF2 is a novel tumor-associated antigen overexpressed in various types of cancers. *Int J Oncol* **44**, 1461–1469 (2014).
97. Wang, Y. *et al.* Tubulin alpha-1b chain was identified as a prognosis and immune biomarker in pan-cancer combing with experimental validation in breast cancer. *Sci Rep* **14**, (2024).
98. Wei, X. Y. *et al.* Characters of KRT80 and its roles in neoplasms diseases. *Cancer Medicine* vol. 12 13991–14003 Preprint at https://doi.org/10.1002/cam4.6040 (2023).
99. He, A., He, S., Li, X. & Zhou, L. ZFAS1: A novel vital oncogenic lncRNA in multiple human cancers. *Cell Proliferation* vol. 52 Preprint at https://doi.org/10.1111/cpr.12513 (2019).

Method details

Cell culture and experimental conditions

MCF7 (HTB-22) and T-47D (HTB-133) cells were grown in RPMI supplemented with 10% (vol/vol) fetal bovine serum FBS (GIBCO), glutamine, and penicillin/streptomycin. HEK-293 (CRL-1573), MDA-MB-468 (HTB-132), MDA-MB-231 (HTB-26), and HeLa cells (CCL-2) were grown in DMEM supplemented with 10% (vol/vol) FBS, glutamine, and penicillin/streptomycin. A-4098 cells (HTB-44) were grown in. EMEM supplemented with 10% FBS, glutamine, and penicillin/streptomycin. SH-SY5Y (CRL-2266) were grown in DMEM/F12 supplemented with 10% FBS, glutamine, and penicillin/streptomycin. HMLE cells were grown in Promocell mammary epithelial cell basal media (C-21010) with added supplements (c-93110), whereas MCF10A were grown on DMEM/F12 supplemented with 5% Horse serum, 20 ng/ml EGF, 0.5 mg/ml Hydrocortisone, 100 ng/ml Cholera toxin, 10 mg/ml Insulin and Pen/Strep. MCF-7 cells were synchronized in the G1 cell cycle stage by culturing in serum-free media for 72 hours followed by replacing the media with 5% FBS media for 5 hours. Cells were grown at 37°C under a humidified atmosphere with 5% CO_2. Cells were routinely authenticated by STR profiling, and tested for mycoplasma, and cell aliquots from early passages were used.

Cells were treated with DMSO (Sigma-Aldrich), dBET6 (Sigma-Aldrich, SML2638; 100 nM), or CDK9i (NVP-2) (MCE, Cat# Y-12214A; 100 nM) for 4 hours. Cells were treated with DMSO or DNA-PKci (NU-7441, Selleckchem; 2 μM) for 5 hours.

Transcription stress score

Our sBLISS, DRIP-seq, TOP1cc, and TOP1 ChIP-seq data used for transcription stress score calculation are already published[13] and are available at GEO: GSE241309. For each dataset, the signal density per gene was first converted to an empirical p-value based on its rank within the distribution of all genes. Higher signal values correspond to lower p-values. These p-values were then transformed into Z-scores using the standard normal inverse cumulative distribution function:

$$Z_k = \Phi^{-1}(1 - p_k)$$

where p_k is the empirical p-value for dataset k, and Φ^{-1} denotes the inverse of the standard normal cumulative distribution function.

The individual Z-scores from all four datasets were combined per gene using the Liptak weighted Z-test (unweighted version):

$$TSS_{raw} = \frac{Z_1 + Z_2 + Z_3 + Z_4}{\sqrt{4}}$$

For better representation, the combined score was further min-max normalized and then stretched by a power transformation to emphasize differences:

$$TSS_{final} = TSS_{scaled}^4$$

Genes not among the top 1000 in one or more parameters were excluded.

Identification of Transcription Stress-Associated Domains using Multivariate HMM

To identify genomic regions simultaneously enriched for markers of transcription stress, we applied a multivariate hidden Markov model (HMM) using the *mhsmm* R package. The hg38 genome was divided into 1 kb non-overlapping bins, and mean signals for TOP1, TOP1cc, R-loops, and DSBs were calculated per bin. Signal intensities were capped at the 0.5th and 99.5th percentiles to minimize the influence of outliers. A two-state multivariate HMM was initialized with uniform state priors and a transition matrix favouring self-transition (0.95 probability). Emissions were modelled using multivariate Gaussian distributions with means offset by ±0.5 standard deviations and shared covariance. The model was trained by maximum likelihood estimation, and Viterbi decoding was used to assign each bin to a state. The state with the highest average signal across all datasets was designated as the "co-marked" state. To refine the identified regions, we selected bins showing high signal (above the 95th percentile) in at least three of the four datasets, merged adjacent bins separated by ≤1 kb, and filtered to retain regions ≥3 kb. [13]

Gene sets compilation and enrichment analysis

745 Gene sets tested for enrichment in this manuscript were defined as follows: SE-regulated
746 genes for each cell line were downloaded from SEdb2.0[31]. Samples IDs used were:
747 Sample_01_0046 for MCF7, Sample_02_1273 for HeLa, Sample_00_0015 for IMR90,
748 Sample_02_0434 for HEK-293, Sample_02_0097 for T-47D, Sample_02_0095 for MDA-MB-
749 468, Sample_02_0219 for MDA-MB-231, Sample_02_0643 for MDA-MB-436,
750 Sample_02_1812 for SHSY5Y, Sample_02_0056 for HMLE, Sample_02_0268 for A4098. A
751 list of annotated oncogenes was downloaded from OncoKB[68]. Highly expressed genes are
752 defined as the top 1000 expressed genes. TE-regulated genes are defined as genes that
753 overlap enhancer regions and are not annotated by an SE. Non-E genes are genes that do not
754 overlap any enhancer element. enrichment was calculated by dividing the number of observed
755 overlapping genes by the theoretical number of genes predicted under a random model. One-
756 tailed (upper-tailed for values > 1, and lower-tailed for values < 1) hypergeometric test using
757 the phyper function in R to determine the statistical significance of the observed enrichment.
758

759 **In-suspension break labelling *in situ* and sequencing (sBLISS)**

760 sBLISS was conducted as previously described[41,69]. In summary, 10^6 cells were fixed in 2%
761 paraformaldehyde in 10% FCS/PBS for 10 minutes at room temperature. The fixation was
762 quenched with 125 mM glycine for 5 minutes at room temperature, followed by another 5 minutes
763 on ice and two washes in ice-cold PBS. Cells were lysed for 60 minutes on ice, and their nuclei
764 were permeabilized for 60 minutes at 37°C. Next, nuclei were rinsed twice with CutSmart Buffer
765 containing 0.1% Triton X-100 (CS/TX100), and double-strand break (DSB) ends were blunted in
766 situ using NEB's Quick Blunting Kit for 60 minutes at room temperature. The blunted nuclei were
767 then washed twice with 1x CS/TX100 before in situ ligation of the sBLISS adapters to the DSB
768 ends. Adaptor ligation was carried out with T4 DNA Ligase for 20-24 hours at 16°C, with BSA
769 and ATP added. Following ligation, the nuclei underwent two washes with 1x CS/TX100, and
770 genomic DNA was extracted using Proteinase K at 55°C for 14-18 hours while shaking at 800
771 rpm. Proteinase K was then heat-inactivated for 10 minutes at 95°C, followed by extraction using
772 Phenol:Chloroform:Isoamyl Alcohol, Chloroform, and ethanol precipitation. The purified DNA
773 was sonicated in 100 µL of ultra-pure water using Covaris M220 for 60 seconds. Sonicated
774 samples were concentrated with AMPure XP beads (Beckman Coulter), and fragment sizes
775 were evaluated using a BioAnalyzer 2100 (Agilent Technologies), targeting a range of 300 bp to

776 800 bp with a peak around 400-600 bp. The sonicated DNA was then in vitro transcribed using
777 the MEGAscript T7 Kit for 14 hours at 37°C. After RNA purification and ligation of the 3'-Illumina
778 adaptors, the RNA underwent reverse transcription. The final library indexing and amplification
779 step was performed with NEBNext® Ultra™ II Q5® Master Mix.
780 sBLISS fastq files are initially de-multiplexed using sample barcodes. Quality control is
781 performed with trim galore to eliminate residual adapters, trim reads to a base quality of at least
782 20, and filter out short reads smaller than 20 bp. Initial and final sample qualities are assessed
783 with fastqc. Quality-processed fastq files are aligned to the GRCh38 assembly with hisat2, then
784 sorted and indexed using samtools. The resulting bam files are de-duplicated with umi-tools,
785 utilizing genomic coordinates and Unique Molecular Identifiers (UMI). Custom Python and R
786 scripts are employed to identify read start positions and convert bam files to bigwig format for
787 downstream analysis. An additional custom R script is used to discard a blacklist of positions,
788 primarily within centromeres.

789 **RNA-seq and GRO-seq data processing**

790 Bulk RNA-seq processed data for untreated MCF7, HeLa, and T-47D cell lines were obtained
791 from ENCODE[70,71], while data for MDA-MB-436 and MDA-MB-468 were downloaded from GEO
792 (GSE212143)[72]. For MCF7, the following datasets were used: ENCFF967AOT, ENCFF328CKZ,
793 ENCFF921PJP, ENCFF930TQB, ENCFF030IUO, ENCFF878CJF, ENCFF057ITQ,
794 ENCFF317XCR, ENCFF985KJE, and ENCFF456OYZ. For HeLa, we used ENCFF750TQI,
795 ENCFF622TLZ, ENCFF689BFU, ENCFF151QTZ, ENCFF262EJQ, ENCFF878CJF,
796 ENCFF964OAC, ENCFF346JDS, and ENCFF638MLA. For T-47D, we used ENCFF234TQB,
797 ENCFF412DJX, ENCFF475SHD, and ENCFF846PLI. TPM values were averaged across
798 replicates for each cell line to determine the expression rank of each gene.
799 RNA-seq data for MCF7 after CDK9i treatment was downloaded from GSE129012 (4h
800 timepoint). And GRO-seq data for MCF7 after E2 treatment was downloaded from GSE27463
801 (40 min timepoint).
802

803 **RT-qPCR and Spike-in RT-qPCR**

804 Total RNA was isolated utilizing TRI reagent (Biolab), following the manufacturer's guidelines for
805 the phenol/chloroform extraction technique. cDNA was synthesized from 1 µg of RNA using the
806 QScript cDNA synthesis kit (Quantabio). The SYBR Green PCR Master Mix (Applied

807 Biosystems) was employed for qRT-PCR. All assays were conducted in triplicate. Experiments
808 were performed to assess the levels of cDNA using primers targeting: Chicken RPL4 (FW,
809 GAGTGACTACAACCTGCCGA; REV, TTGGCGTATGGGTTCAGCTT), SIK1 (FW,
810 CTCCGGGTGGGTTTTTACGAC; REV, CTGCGTTTTGGTGACTCGATG), ID3 (FW,
811 GAGAGGCACTCAGCTTAGCC; REV, TCCTTTTGTCGTTGGAGATGAC), XBP1 (FW,
812 CCCTCCAGAACATCTCCCCAT; REV, ACATGACTGGGTCCAAGTTGT), CCND1 (FW,
813 GCTGCGAAGTGGAAACCATC; REV, CCTCCTTCTGCACACATTTGAA), HES1 (FW,
814 CCTGTCATCCCCGTCTACAC; REV, CACATGGAGTCCGCCGTAA), FOS (FW,
815 TCCCATCGGTCCACTAGGTTT; REV, AGGGCTGCACTGAGTTCTTTG), JUNB (FW,
816 ACGACTCATACACAGCTACGG; REV, GCTCGGTTTCAGGAGTTTGTAGT), TFF1 (FW,
817 CCCTCCCAGTGTGCAAATAAG; REV, GAACGGTGTCGTCGAAACAG), UBC (FW,
818 ATTTGGGTCGCGGTTCTTG; REV, TGCCTTGACATTCTCGATGGT), HPRT (FW,
819 TGACACTGGCAAAACAATGCA; REV, GGTCCTTTTCACCAGCAAGCT), MYC (FW,
820 ATGCCCCTCAACGTGAACTTC; REV, CGCAACATAGGATGGAGAGCA), XRCC4 (FW,
821 ATGTTGGTGAACTGAGAAAAGC; REV, GCAATGGTGTCCAAGCAATAAC), GREB1 (FW,
822 TGGTCCGTAATGCACAAGGG; REV, CTGCGTTTAGTGAGGGGTGA), GATA3 (FW,
823 GCCTCTGCTTCATGGATCCC; REV, CACACTCCCTGCCTTCTGTG), ACTB (FW,
824 TTTTGGCTATACCCTACTGGCA; REV, CTGCACAGTCGTCAGCATATC).
825 In RT-qPCR experiments, Ct values were normalized to HPRT levels. For spike-in RT-qPCR,
826 chicken RNA was added in proportion to the number of cells in each sample, and Ct values were
827 normalized to chicken RPL4.
828

829 **γH2AX ChIP-seq**

830 MCF7 cells (~10^6) were crosslinked with 1% formaldehyde (methanol free, Thermo Scientific
831 28906) for 10 min at room temperature and quenched with glycine, 125 mM final concentration.
832 Fixed cells were washed twice in PBS and incubated in sonication buffer (0.5M NaCl, 0.5%SDS
833 RIPA buffer containing PMSF, protease and phosphatase inhibitors) for 30min on ice. Cells were
834 sonicated using Covaris M220 for 10min to produce chromatin fragments of ~200-300 bp. The
835 sheared chromatin was centrifuged 10min at maximum speed. From the supernatant, 50ul were
836 saved as input DNA and the rest was diluted in RIPA buffer without NaCl and SDS. The
837 chromatin was immunoprecipitated by incubation with 5 µg of gamma-H2AX (phospho S139;

838 Abcam, AB2893) antibody pre mixed with 100ul Dynabeads Protein G (ThermoFisher, 10004D)
839 and incubated over-night at 4°C with rotation. Immunoprecipitates were washed twice with RIPA
840 buffer + 150mM NaCl, twice with RIPA buffer + 300 mM NaCl, and twice with TE buffer. The
841 chromatin was eluted from the beads with 200ul of direct elution buffer (10mM Tris pH 8, 0.3M
842 NaCl, 5mM EDTA, 0.5% SDS) and incubated overnight at 65°C to reverse the cross-linking.
843 Samples were treated with RNAse for 1h at 37 °C and with proteinase K at 55 °C for 2 h. DNA
844 was cleaned up by QIAquick PCR purification column (Qiagen), according to the manufacturer's
845 instructions, and eluted in 30 µl of elution buffer. The ChIPed and the Input DNA were used for
846 real-time PCR or to prepare libraries by Truseq (Ilumina) and sequenced in Nextseq (Illumina).
847 ChIPed and input fastq files are quality controlled with trim galore to remove residual adapters,
848 trim reads to base quality of at least 20 and filter out short reads with size smaller than 20 bp.
849 Initial and final sample qualities are evaluated with fastqc. The quality processed fastq files are
850 aligned to GRCh38 genome with hisat2 followed by sorting and indexing with samtools. The
851 utility bamCompare of deepTools is applied on the bam files to create a peak bigwig file. This
852 operation is set to ignore duplicates, discard a blacklist of regions, normalize by RPKM, and
853 compare by ratio in tiles of 50 bp. γH2AX ChIP-seq data for HeLa and IMR90 were downloaded
854 from GEO: GSM4564685 and GEO: GSM2067927, respectively. γ H2AX density per gene was
855 computed using bigWigAverageOverBed, which calculates the average signal from the γ H2AX
856 bigWig track over annotated gene intervals.
857
858 **Statistical modelling and residual-based classification**
859 For multivariate logistic regression, genes were categorized based on whether they overlapped
860 transcription stress sites identified by Hidden Markov Modeling (HMM). Logistic regression was
861 performed using the glm function in R (family = binomial).
862 Multivariate Least Absolute Shrinkage and Selection Operator (LASSO) regression was
863 conducted using the glmnet package in R, with 10-fold cross-validation (cv.glmnet) to select the
864 optimal regularization parameter (λ). The model was trained to classify genes based on
865 transcription stress site overlap or DSB enrichment. Odds ratios were derived from the final
866 model coefficients to assess the relative predictive power of each feature.
867 Generalized additive model (GAM) was applied to model the relationship between each variable
868 and γH2AX across all genes in (Fig. 5A), and for genes classified into their overlapping chromatin

869 states in (Fig. 5H). The model was constructed using the gam function in the mgcv package in
870 R.
871 To dissect the effects of absolute DSBs from DSB turnover on mutation frequency, we modelled
872 the relationship between break density and transcription stress by nonlinear least squares (NLS)
873 regression using the following equation:

874 $$gH2AX\ density = a \times break\ density^{b}$$

875 The model was fitted using the nls function in R, with initial parameter estimates
876 for “a” and “b” set to 0.1. After fitting the model, residuals were extracted and used to classify
877 genes into three response categories based on predefined thresholds: genes with residuals > 0
878 were classified as 'Over-responsive,' genes with residuals < 0 as 'Under-responsive' and genes
879 between -0.3 and 0.3 as 'Moderately responsive'.

880 **Classifying individual DSB sites based on γH2AX**

881 Each break site was assigned a γH2AX coverage level (γH2AX rank). To ensure a comparable
882 distribution of break scores (i.e., the number of DSBs at the same site) between the two groups,
883 break sites were classified into four categories: those with a break score of 1 (the large majority),
884 those with a break score between 2 and 4, those with a break score between 5 and 10, and
885 those with a break score between 11 and 50. Within each of the four categories, 1% of the top
886 and bottom γH2AX-ranked breaks are selected. All categories are then re-combined so that the
887 top set of breaks and the bottom set of breaks exhibit a similar break score distribution. Mapping
888 to chromatin states is then computed and plotted for these sets.
889

890 **Chromatin states**

891 Chromatin states bed files were downloaded from ENCODE[70,71]. For MCF7, ENCFF631HCR file
892 was used. For HeLa, ENCFF555HXM file was used.
893

894 **Transient transfection**

895 Transient transfection of siXRCC4 (Dharmacon; SMARTpool siRNA, Catalog ID: L-004494-00-
896 0005) and siSc (Dharmacon; D-00181010) was carried out using Lipofectamine (ThermoFisher).
897 On the day of transfection, cells were seeded to achieve a confluency of 60-70%. The
898 transfection solution was prepared by mixing 1.5 ml of serum and antibiotic-free RPMI with 30

899 μL of Lipofectamine and incubating for 5 minutes before adding a mixture of 1.5 ml RPMI and
900 0.6 nmoles of siRNA, followed by a 20-minute incubation at room temperature. This mixture was
901 then added to cells in a 10 cm plate cultured in antibiotic and serum-free media. The media was
902 replaced with fully supplemented media after 5-6 hours, and the cells were incubated in the
903 incubator for 48 hours.
904
905 **HiC data analysis**
906 Normalized Hi-C contact matrices for MCF7 and MCF10A cells were obtained from GEO
907 (GSE66733).
908 Genomic bins were intersected with all genes, and each bin was assigned to a single gene. Bins
909 overlapping multiple genes were excluded from the analysis.
910 In inter- and intra-group metrics calculation, genes were stratified into nine groups based on their
911 break density, excluding genes with zero breaks. Intra-group and inter-group interaction scores
912 were computed as follows: Intra-group interactions: For each gene, interaction scores with all
913 other genes within the same group were summed and normalized by the total number of genes
914 in that group. Inter-group interactions: For each gene, interaction scores with all genes outside
915 its group were summed and normalized by the total number of genes in all other groups. Genes,
916 along with their interaction scores, were ordered based on break density. The data was
917 visualized using loess smoothing.
918
919 **COSMIC mutations data analysis**
920 The file Cosmic_NonCodingVariants_Tsv_v101_GRCh37.tar, containing all noncoding variants,
921 was downloaded from COSMIC. This file was first filtered for breast cancer samples and
922 separated based on the type of mutation into SNVs, deletions, and insertions mutation files.
923 These files were then converted into genomic ranges objects and filtered to contain only
924 mutations within introns. To calculate mutation density, mutations per gene were counted and
925 normalized by the sum of intron length.
926
927
928
929

Acknowledgments

We sincerely thank the members of the Aqeilan lab for their valuable discussions and insights. We are deeply grateful to Prof. Sheera Adar and Dr. Yotam Drier for their thoughtful guidance and support as members of the PhD committee. We also appreciate the invaluable assistance of Dr. Abed Nasereddin and Dr. Idit Shiff from the Core Research Facility of the Hebrew University-Hadassah Medical School. Additionally, we extend our sincere gratitude to Prof. Yinon Ben Neriah for generously providing dBET6 and CDK9i. This study was supported by a grant from the Israel Science Foundation (ISF) [No. 1056/21]. We also acknowledge the support of the Carole and Andrew Harper Diversity Scholarship Program to O.H. and D.S, Science Training Encouraging Peace (STEP) Fellowship to D.S. and S.O. and the VATAT PhD scholarship to O.H.

Author contributions

O.H. and R.I.A. conceived and designed the study. O.H. and J.M. performed computational analysis. S.O.F performed sBLISS on HMLE, MCF10A, MCF7, T47D, MDA-MB-436 and MDA-MB-468 cell lines. D.S. performed sBLISS on MCF7 cells synchronized in G1, SH-SY5Y cells, and NSCs. O.H. performed all other experiments. O.H. and R.I.A wrote the manuscript.

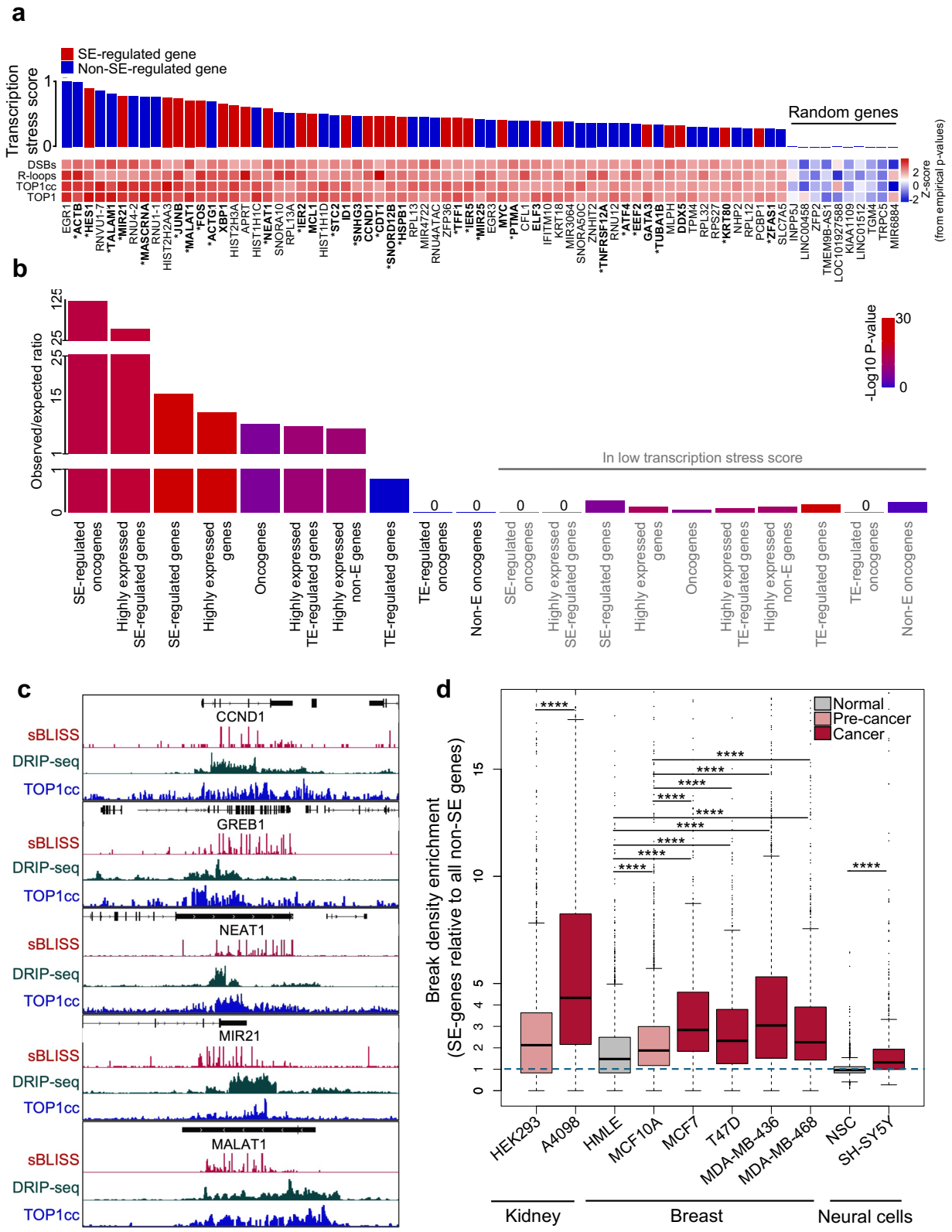


Fig. 1: Super-enhancer-regulated oncogenes are major transcription stress sites.

954 **a,** Transcription stress sites ranked by their scores based on four transcription stress
955 parameters: TOP1, R-loops, TOP1cc, and DSBs in MCF7 cells (see methods). Genes with
956 known oncogenic function in OncoKB are in bold and from other resources in bold with an
957 asterisk (see table 1). **b,** The ratio of observed to expected numbers of genes in transcription
958 stress sites. P-values were calculated using hypergeometric testing (upper-tailed for enrichment
959 in the top score and lower-tailed for the bottom score). **c,** Genome browser snapshots depict
960 DSBs, TOP1cc, and R-loops at selected transcription stress sites. **d,** A boxplot illustrates DSBs
961 enrichment at the SE-regulated genes over other genes in each cell line (downloaded from
962 SEdb[2.0 31]). Enrichment was calculated by dividing the break density of each SE-regulated gene
963 by the median break density of non-SE-regulated genes and the average values for replicates
964 were plotted. In the boxplot, the center line represents the median; the box limits indicate the
965 upper and lower quartiles; the whiskers denote 1.5X the interquartile range, while the dots
966 represent outliers. P-values were computed using the Mann-Whitney U test, ****$p < 0.0001$.

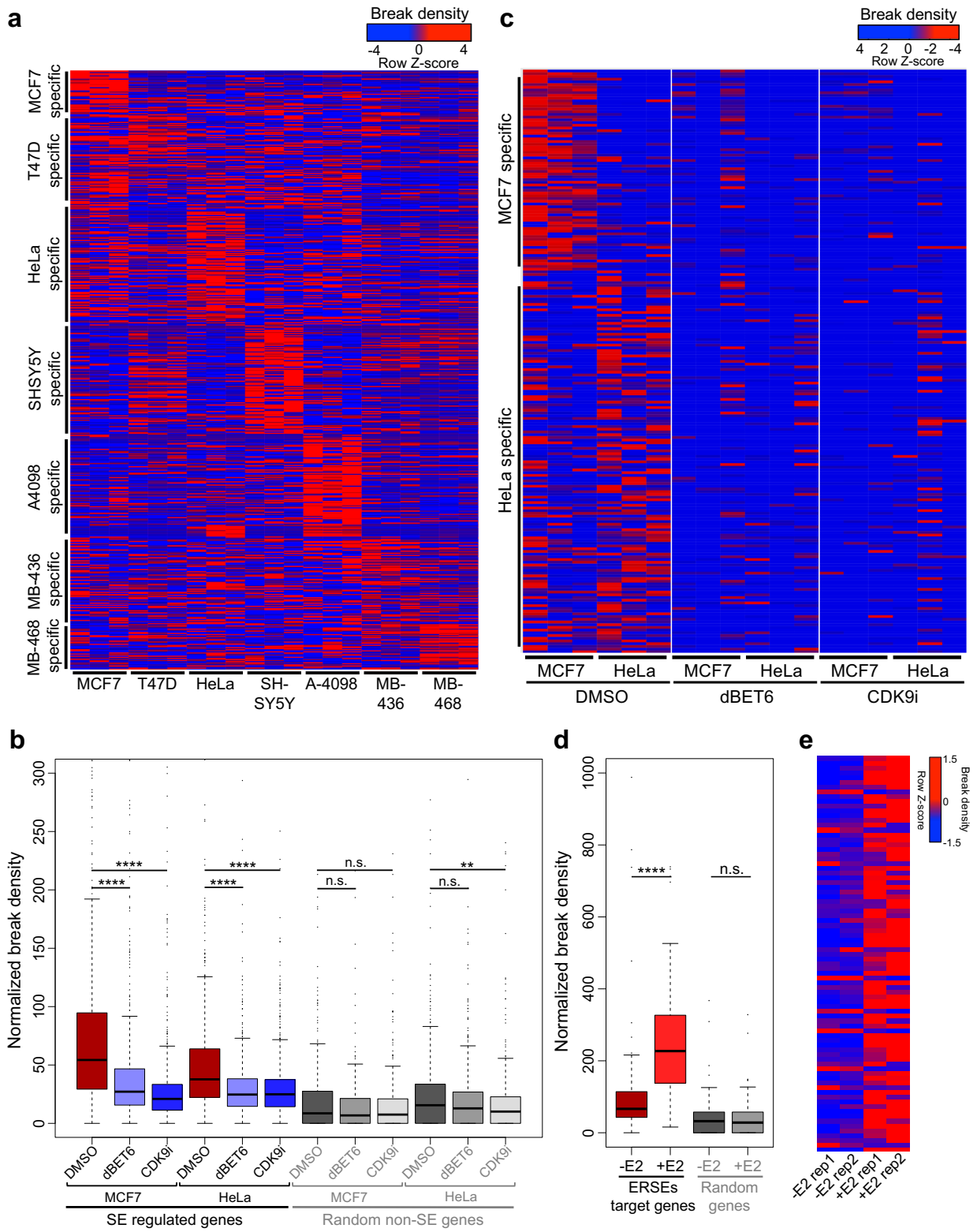


Fig. 2: Super-enhancer landscape shapes transcription stress sites in cancer.
a, Heatmap showing the normalized break density row Z-score for each cell line's unique SE-regulated genes. For each cell line, SE-regulated genes were identified, common SE-regulated

genes from other cell lines were excluded, and low to moderately expressed genes as well as genes with low break counts were filtered out. Each column corresponds to a biological replicate. Columns were normalized by median scaling. **b**, Boxplot illustrating the normalized break density of cells treated with DMSO (control), dBET6, or CDK9i (100nM for 4 hours each) for SE-regulated genes in each cell line or for randomly selected genes. values are averaged for three biological replicates. **c**, Heatmap depicting the normalized break density row Z-score for each cell line's unique SE-regulated genes following treatment with DMSO (control), dBET6, or CDK9i. Each column corresponds to a biological replicate. Each sample was normalized by downsizing to 270000 random breaks. **d**, Boxplot demonstrating the normalized break density of cells treated with vehicle (control) or estradiol (E2, 100nM for 1 hour) for significantly affected ($p < 0.05$) estrogen receptor-occupied SE (ERSE) target genes or for randomly selected genes. **e**, Heatmap representing the normalized break density row Z-score of significantly affected ERSE target genes. Boxplot features include: center line representing median; box limits denoting upper and lower quartiles; whiskers extending to 1.5x the interquartile range; and dots indicating outliers. P-values were computed using the Mann-Whitney U test, with n.s. indicating $p > 0.05$, ** indicating $p < 0.01$, and **** indicating $p < 0.0001$.

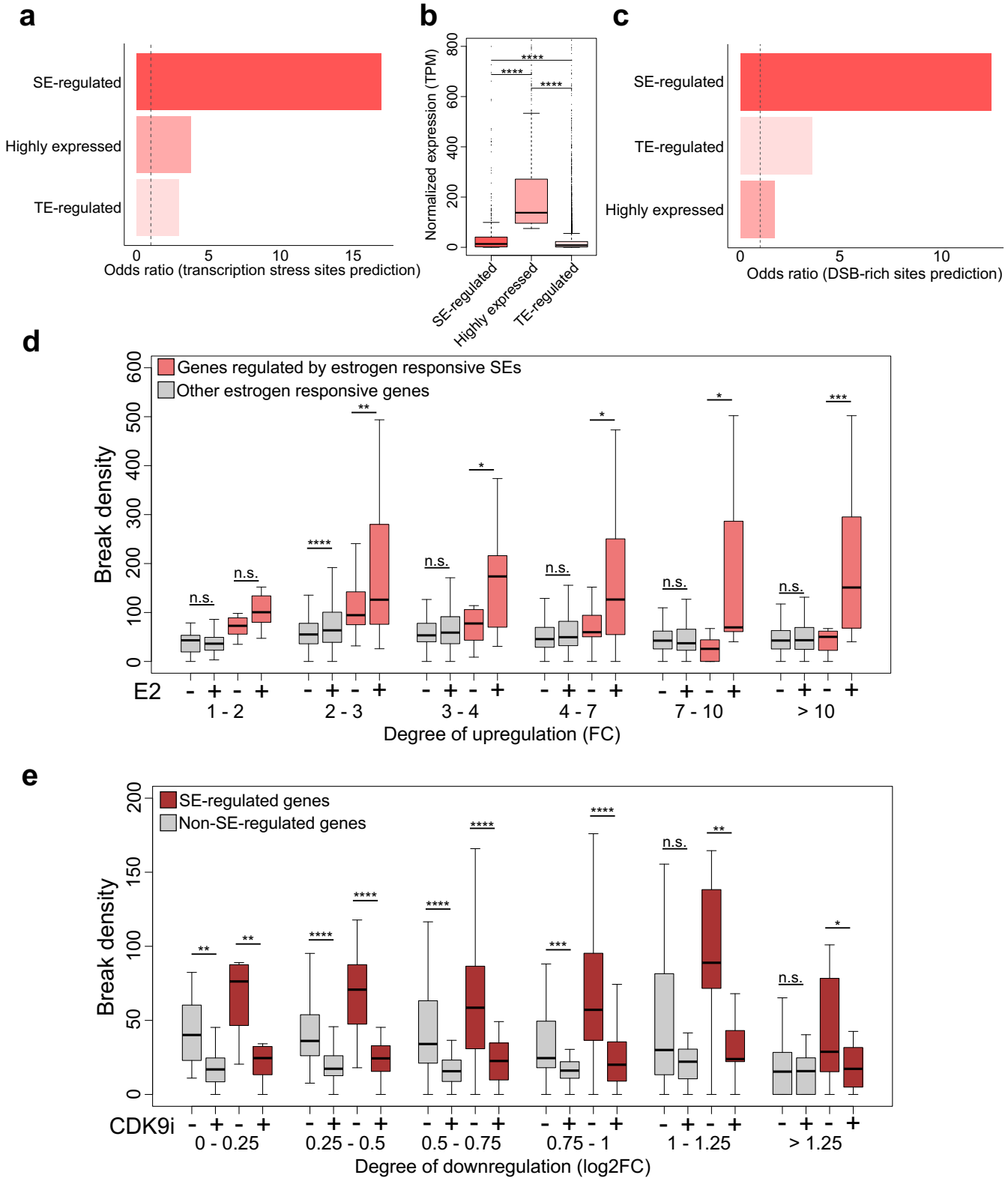


Figure 3. Specificity of transcription stress at super-enhancer (SE)–regulated genes. **a,** Multivariate LASSO regression identifying predictors of transcription stress site locations in 100kb genomic bins. Odds ratios are shown. **b,** Boxplot showing normalized expression (TPM)

991 of SE-regulated, highly expressed, and TE-regulated genes. Despite lower expression. **c,**
992 LASSO regression for prediction of DSB-rich genomic sites. the genome was binned in 10kb
993 bins, and the top 5% bind in break density were assigned as DSB-rich. **d,** Boxplot showing the
994 break density in estrogen-responsive genes upon E2 treatment, grouped by degree of
995 transcriptional upregulation (fold change, FC). Non-SE genes were randomly downsampled to
996 match SE gene numbers for comparable P-values. **e,** Boxplot showing the break density in
997 estrogen-responsive genes following CDK9 inhibition (CDK9i), grouped by degree of
998 transcriptional downregulation (log2 fold change). Non-SE genes were randomly downsampled
999 to match SE gene numbers for comparable P-values. Boxplot features include: center line
1000 representing the median; box limits denoting upper and lower quartiles; whiskers extending to
1001 1.5x the interquartile range; and dots indicating outliers. P-values in boxplots were calculated
1002 using the Mann-Whitney U test in b, and Wilcoxon paired test in D and E. n.s. $p > 0.05$, *$p < 0.05$,
1003 **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

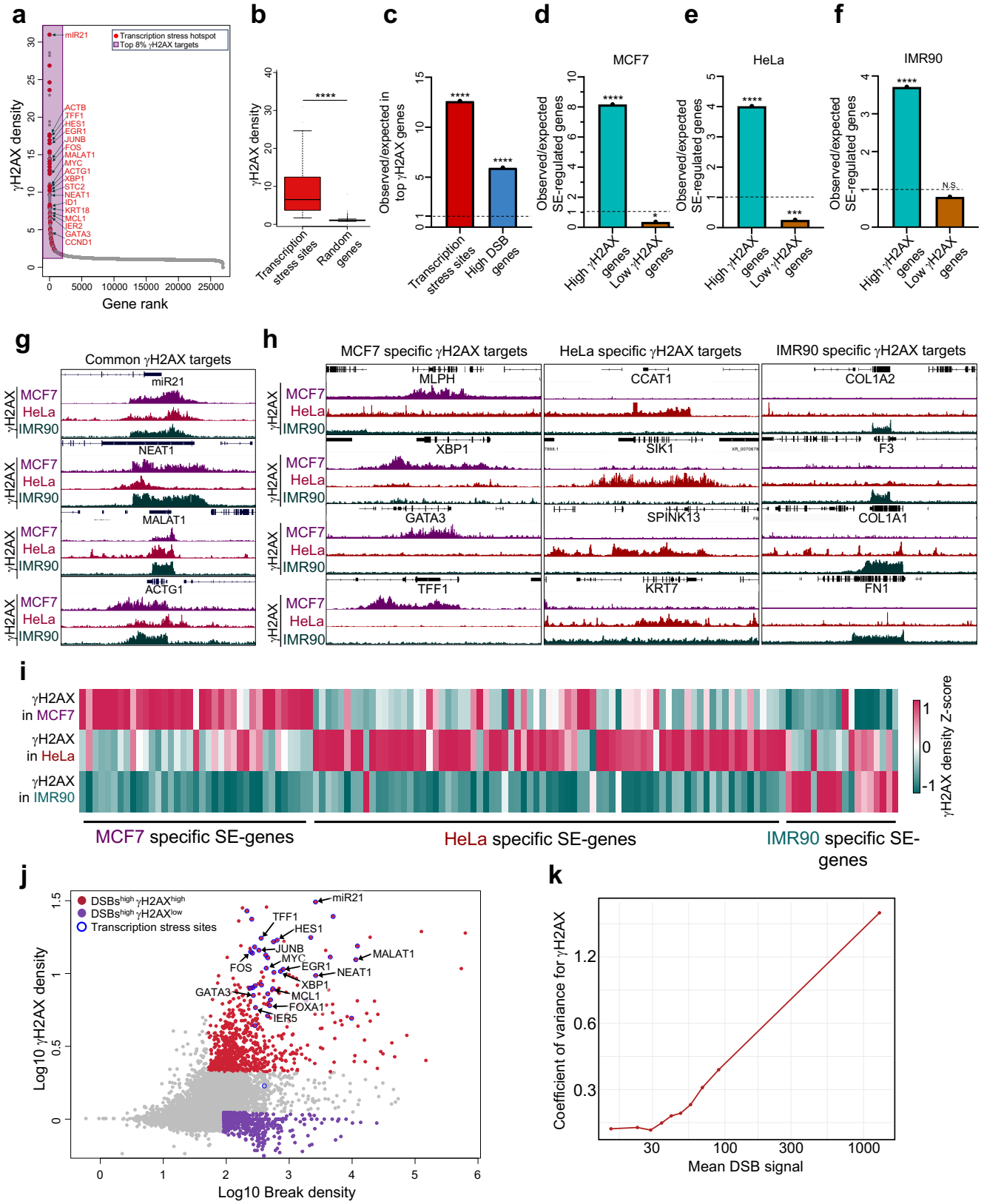
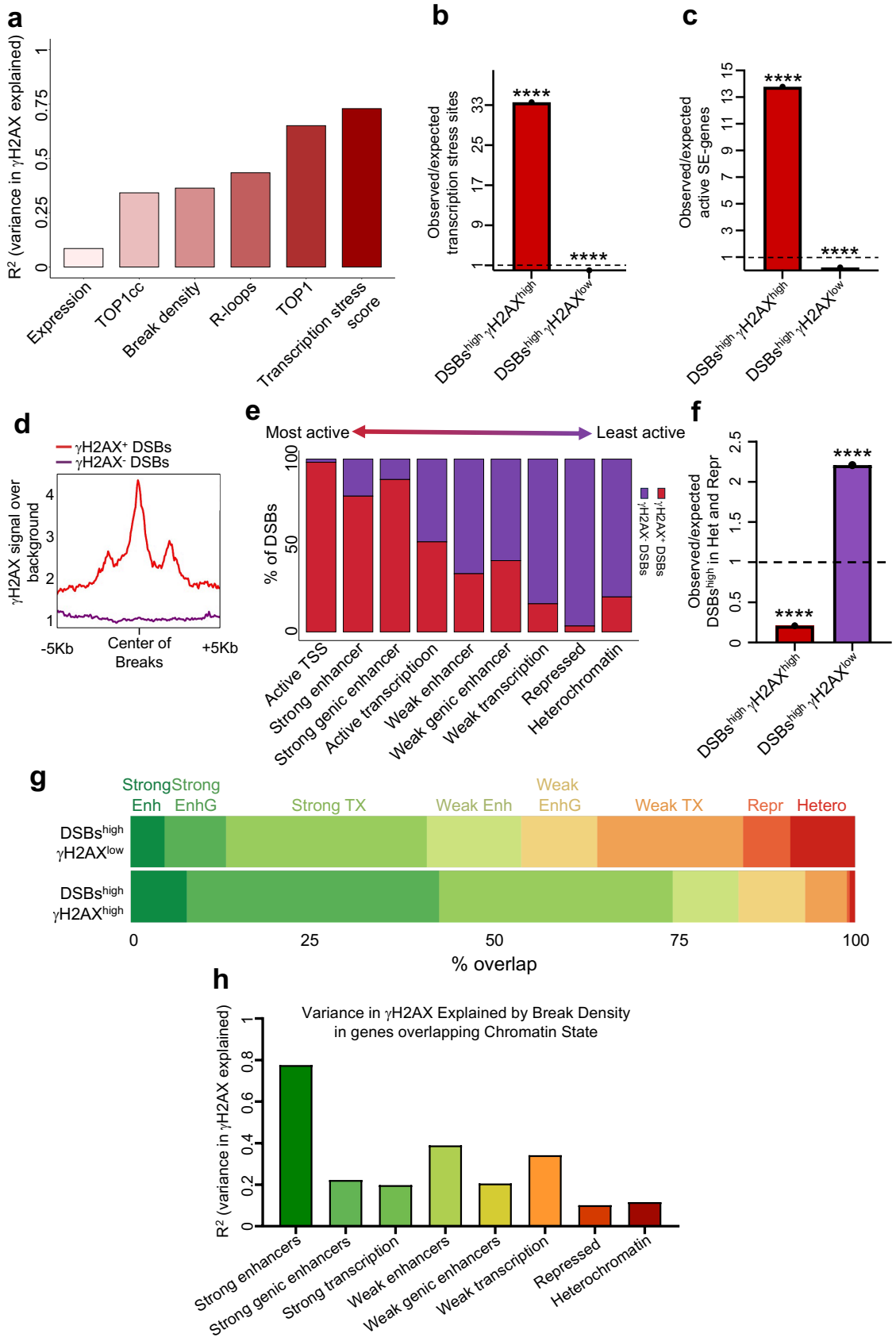


Fig. 4: γH2AX marks transcription stress sites.

1006 **a**, Ranked plot of genes in MCF7 cells based on γH2AX density. γH2AX density per gene was
1007 computed using bigWigAverageOverBed, which calculates the average signal from the γH2AX
1008 bigWig track over annotated gene intervals. **b**, Boxplot showing γH2AX density for transcription
1009 stress sites and randomly selected genes. **c**, The ratio of the observed to the expected number
1010 of transcription stress sites or top 1000 break-prone genes in MCF7 among the top γH2AX
1011 enriched genes. **d-f**, The ratio of observed to expected numbers of SE-regulated genes in the
1012 top and bottom 300 genes enriched in γH2AX for each cell line. P-values were determined using
1013 the hypergeometric test: n.s. $p > 0.05$, *$p < 0.05$, ***$p < 0.001$, ****$p < 0.0001$. **g,** Genome
1014 browser snapshots of representative common γH2AX targets across MCF7, HeLa, and IMR90
1015 cell lines. **h**, Genome browser snapshots of representative unique γH2AX targets in MCF7,
1016 HeLa, and IMR90 cell lines. **i**, Heatmap displaying the γH2AX density column Z-score for each
1017 cell line's unique SE-regulated genes. For each cell line, SE-regulated genes were identified,
1018 common SE-regulated genes with other cell lines, and genes with low γH2AX counts were
1019 filtered out. **j**, Scatter plot showing the correlation between break density and γH2AX density for
1020 each gene. **k,** Coefficient of variation (CV) of γH2AX density as a function of mean DNA double-
1021 strand break (DSB) signal. Genes were binned into deciles based on their DSB signal intensity
1022 (log-transformed), and for each bin, the mean γH2AX density, standard deviation, and CV
1023 (standard deviation divided by mean) were calculated. The x-axis shows the mean DSB signal
1024 per bin on a log scale, and the y-axis shows the CV of γH2AX density within each bin. Boxplot
1025 features include: center line representing the median; box limits denoting upper and lower
1026 quartiles; whiskers extending to 1.5x the interquartile range; and dots indicating outliers. P-
1027 values in boxplots were calculated using the Mann-Whitney U test, ****$p < 0.0001$.


a
R² (variance in γH2AX explained)
0
0.25
0.5
0.75
1
Expression
TOP1cc
Break density
R-loops
TOP1
Transcription stress score
b
Observed/expected transcription stress sites
1
9
17
25
33
****
****
DSBs^high γH2AX^high
DSBs^high γH2AX^low
c
Observed/expected active SE-genes
1
3
5
7
9
11
13
15
****
****
DSBs^high γH2AX^high
DSBs^high γH2AX^low
d
γH2AX⁺ DSBs
γH2AX⁻ DSBs
γH2AX signal over background
1
2
3
4
-5Kb
Center of Breaks
+5Kb
e
Most active
Least active
% of DSBs
0
50
100
γH2AX⁻ DSBs
γH2AX⁺ DSBs
Active TSS
Strong enhancer
Strong genic enhancer
Active transcriptioon
Weak enhancer
Weak genic enhancer
Weak transcription
Repressed
Heterochromatin
f
Observed/expected DSBs^high in Het and Repr
0
0.5
1
1.5
2
2.5
****
****
DSBs^high γH2AX^high
DSBs^high γH2AX^low
g
Strong Enh
Strong EnhG
Strong TX
Weak Enh
Weak EnhG
Weak TX
Repr
Hetero
DSBs^high γH2AX^low
DSBs^high γH2AX^high
0
25
50
75
100
% overlap
h
Variance in γH2AX Explained by Break Density in genes overlapping Chromatin State
R² (variance in γH2AX explained)
0
0.2
0.4
0.6
0.8
1
Strong enhancers
Strong genic enhancers
Strong transcription
Weak enhancers
Weak genic enhancers
Weak transcription
Repressed
Heterochromatin

1028

1029 **Fig. 5: γH2AX preferentially marks endogenous DSBs of active chromatin**.
1030 **a**, Barplot showing R^2 values for the different variables tested regarding their effect on γH2AX
1031 variance across all genes. **b-c**. The ratio of observed to expected number of transcription stress
1032 sites (**b**) and active SE-regulated genes (**c**). **d**, γH2AX signal at the two defined sets of DSBs. **e**.
1033 The percentage of DSBs enriched with γH2AX ($\gamma H2AX^{high}$ DSBs) or lacking γH2AX ($\gamma H2AX^{low}$
1034 DSBs) in each chromatin state. Chromatin states were ordered based on what is typically known
1035 about their transcriptional activity[73]. **f**, The ratio of observed to expected numbers of high DSB
1036 genes overlapping heterochromatin and repressed regions in both gene groups in MCF7 cells.
1037 **g**. The percentage of overlaps between each gene group and chromatin states in MCF7. For
1038 each group, the number of overlaps was counted for each chromatin state and then divided by
1039 the total number of overlaps. **h**. A barplot showing R^2 values for the effect of DSBs on γH2AX
1040 variance for genes overlapping different chromatin states. P-values were determined using the
1041 hypergeometric test: ****$p < 0.0001$.

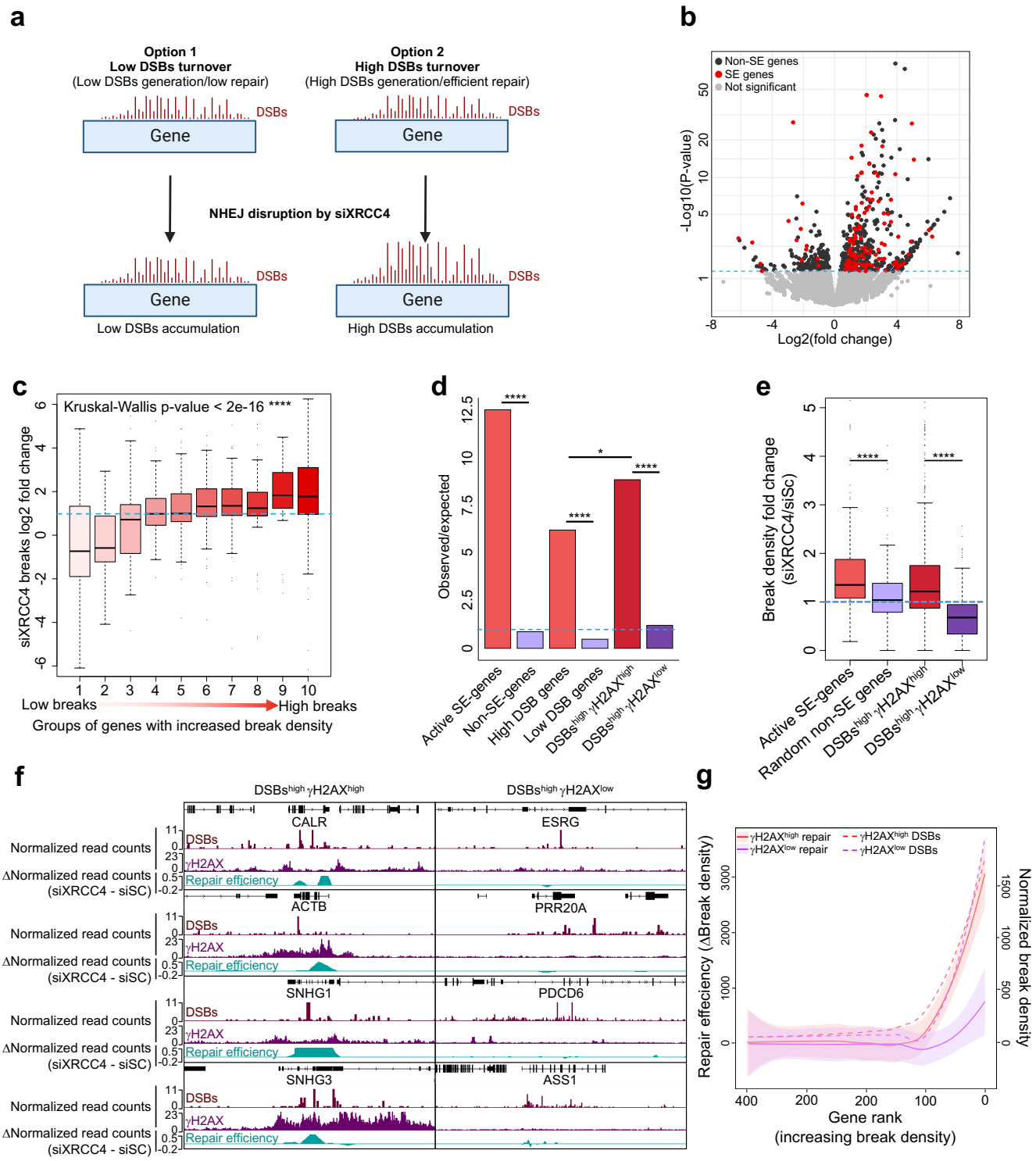


Fig. 6: Genes marked with γH2AX show high DSB turnover and repair efficiency. a. Schematic representation of the rationale behind the experiment to estimate the rate of endogenous DSB repair and turnover. **b**. Volcano plot depicting the results of the differential expression analysis of DSBs following XRCC4 KD, assessed using the DESeq2 algorithm. The experiment was performed on two biological replicates per condition. **c**. Boxplot illustrating the

log2FC of break density upon XRCC4 KD for significantly affected genes stratified according to their baseline break density into ten groups. **d**. The ratio of observed to expected number of genes in our defined gene sets. P-values were calculated using Fischer exact test. **e**. Boxplot of break density fold change. P-values were calculated using the Mann-Whitney U test. **f**. Genome browser snapshots of DSBs^{high} γH2AX^{high} and DSBs^{high} γH2AX^{low} gene examples, showing DSBs, γH2AX, and repair efficiency. Repair efficiency was estimated by pooling replicates, tiling the genome into 100bp tiles, subtracting the break score of each tile in the siXRCC4 sample from siSc (control), and smoothing the segments through sliding window averaging. **g**. Line plot for repair efficiency (log2FC after XRCC4 KD) for DSBs^{high} γH2AX^{high} versus DSBs^{high} γH2AX^{low} genes ordered by their break density. Lines were generated using LOESS smoothing. Boxplot: center line indicates median; box limits represent upper and lower quartiles; whiskers extend to 1.5x interquartile range; dots indicate outliers. *$p < 0.05$, ****$p < 0.0001$.

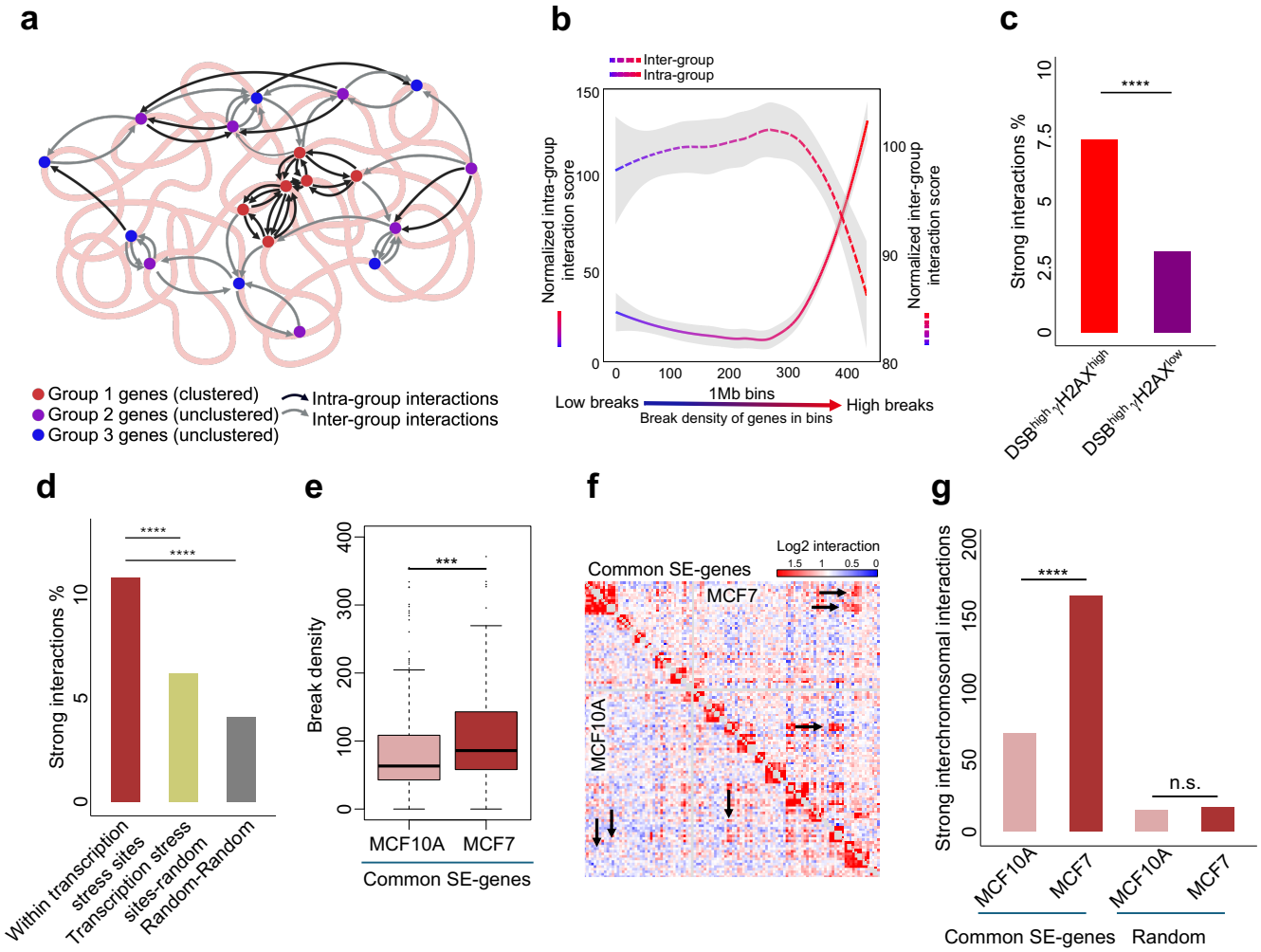


Fig. 7: Efficiently repaired transcriptional DSBs cluster in the 3D genome.
a. Schematic illustrating inter- and intra-group interactions and their use to determine the clustering of genes with a common characteristic (high DSB density in this instance). **b**. Line plot depicting inter-group (dashed line) and intra-group (solid line) genomic bins arranged according to the break density of their overlapping genes. Lines were created using LOESS smoothing, with shaded areas representing the 95% confidence intervals. **c**. Percentage of strong interactions within genomic bins containing DSBs^{high} γH2AX^{high} compared to bins with DSBs^{high} γH2AX^{low} genes. Strong interactions are defined as a log2 interaction score > 2. The percentage is calculated by dividing the number of strong interactions by the total number of interactions. P-values were calculated by Chi-squared test. **d**. Percentage of strong interactions calculated as in (**c**). **e**. Boxplot reflecting break density of common SE-regulated genes between MCF7 and MCF10A cell lines. P-values were calculated by Mann-Whitney U test. **f**. Heatmap showing Log2 interaction scores for genomic bins containing common SE-regulated genes from both MCF7 and MCF10A cell lines. Arrows indicate distinct sites of variation in interactions. **g**. Count of interchromosomal strong interactions shown in the heatmap from (**f**). Random bins serve as a reference. Boxplot: center line indicates median; box limits represent upper and lower quartiles; whiskers extend to 1.5x interquartile range; dots indicate outliers. P-values were calculated using chi-square test in (**c**-**d**) and (**g**), and Wilcoxon paired test in (e). ***p < 0.001, ****p < 0.0001.

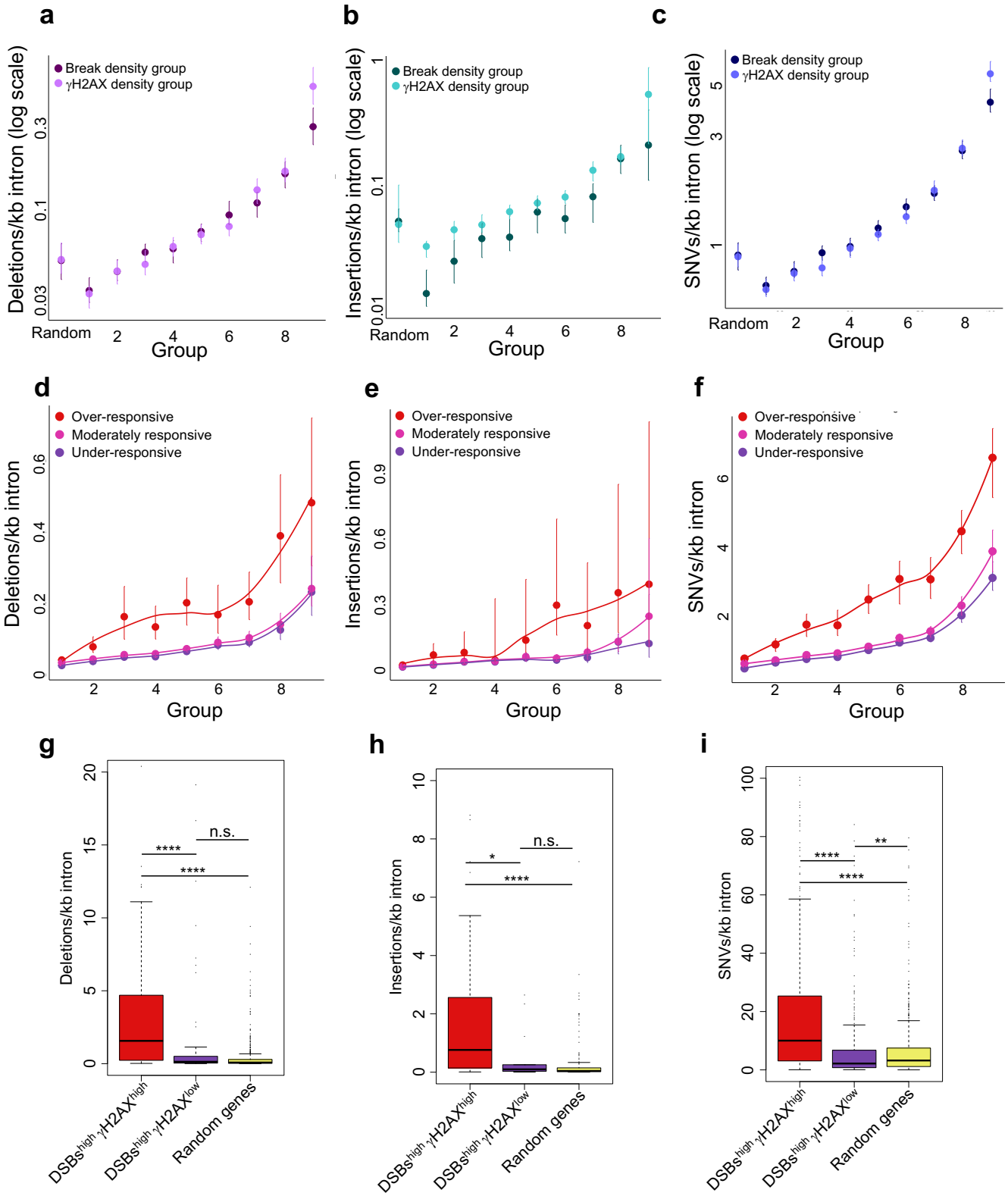
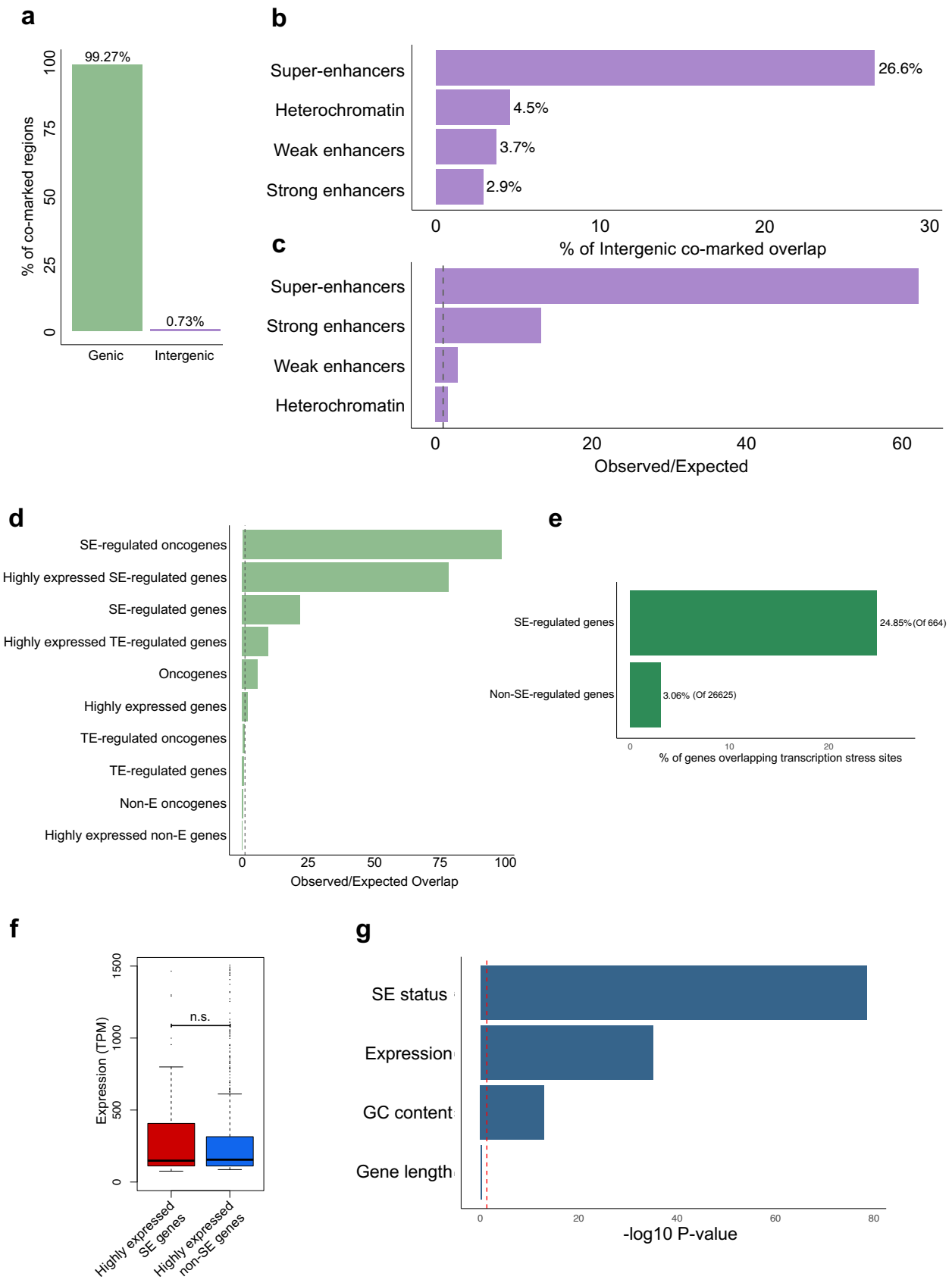


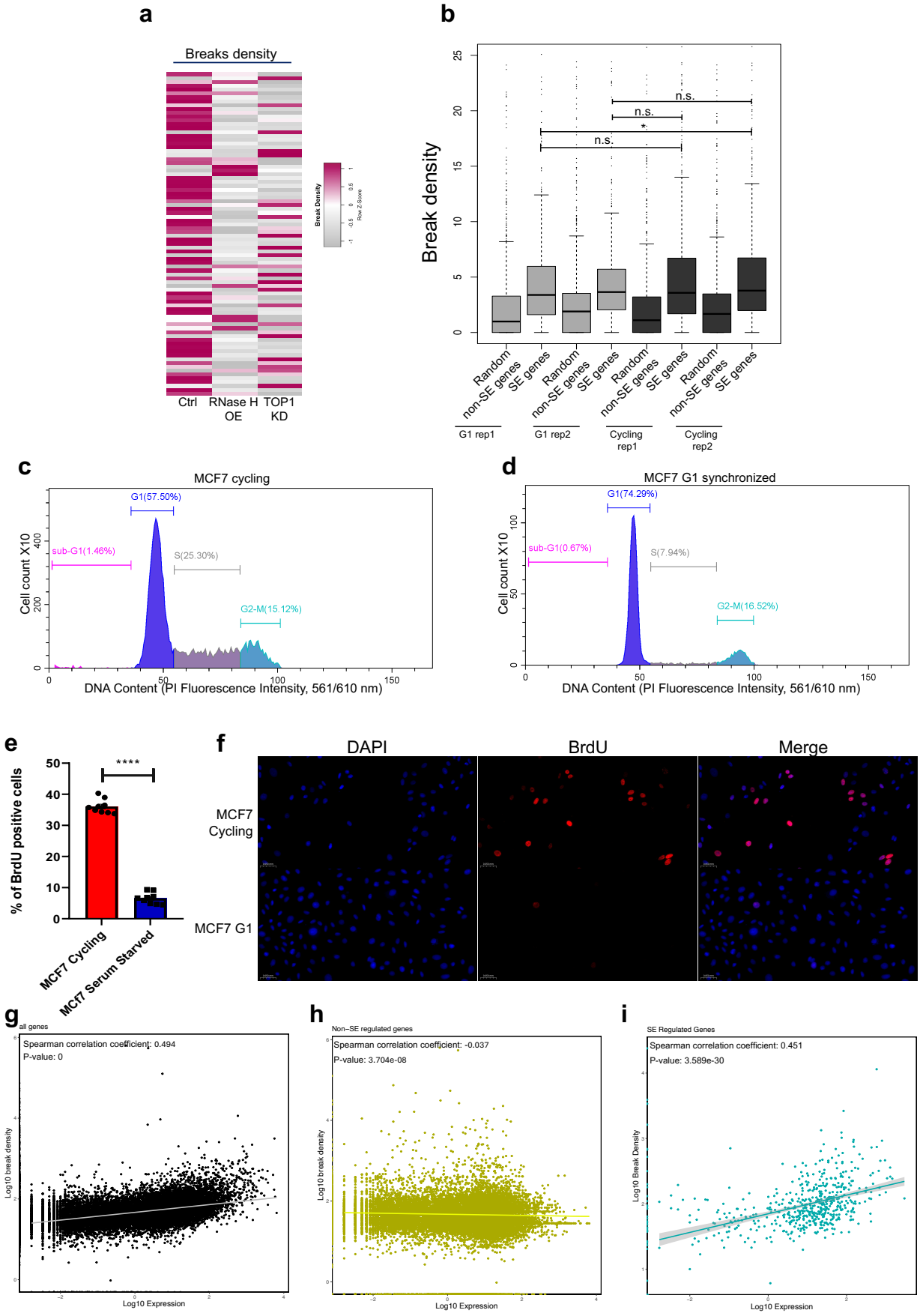
Fig. 8: transcriptional DSBs with high turnover are especially susceptible to mutations.

1089 **a**-**c**. Intronic deletion density (**a**), insertion density (**b**), and SNV density (**c**) for genes categorized
1090 by their break or γH2AX density (excluding zero break genes). Random genes are included for
1091 reference. **d**-**f**, Rate of intronic mutations across genes categorized by their break density and
1092 further divided into three groups based on their γH2AX response to DSBs. **g-i**. Boxplot of intronic
1093 deletion density (**g**), insertion density (**h**), and SNV density (**i**) for $DSBs^{high}$ $\gamma H2AX^{high}$ vs. $DSBs^{high}$
1094 $\gamma H2AX^{low}$ vs. randomly selected genes.


a
% of co-marked regions
99.27%
0.73%
Genic
Intergenic
b
Super-enhancers
Heterochromatin
Weak enhancers
Strong enhancers
26.6%
4.5%
3.7%
2.9%
% of Intergenic co-marked overlap
c
Super-enhancers
Strong enhancers
Weak enhancers
Heterochromatin
Observed/Expected
d
SE-regulated oncogenes
Highly expressed SE-regulated genes
SE-regulated genes
Highly expressed TE-regulated genes
Oncogenes
Highly expressed genes
TE-regulated oncogenes
TE-regulated genes
Non-E oncogenes
Highly expressed non-E genes
Observed/Expected Overlap
e
SE-regulated genes
Non-SE-regulated genes
24.85% (Of 664)
3.06% (Of 26625)
% of genes overlapping transcription stress sites
f
Expression (TPM)
n.s.
Highly expressed SE genes
Highly expressed non-SE genes
g
SE status
Expression
GC content
Gene length
-log10 P-value

1095

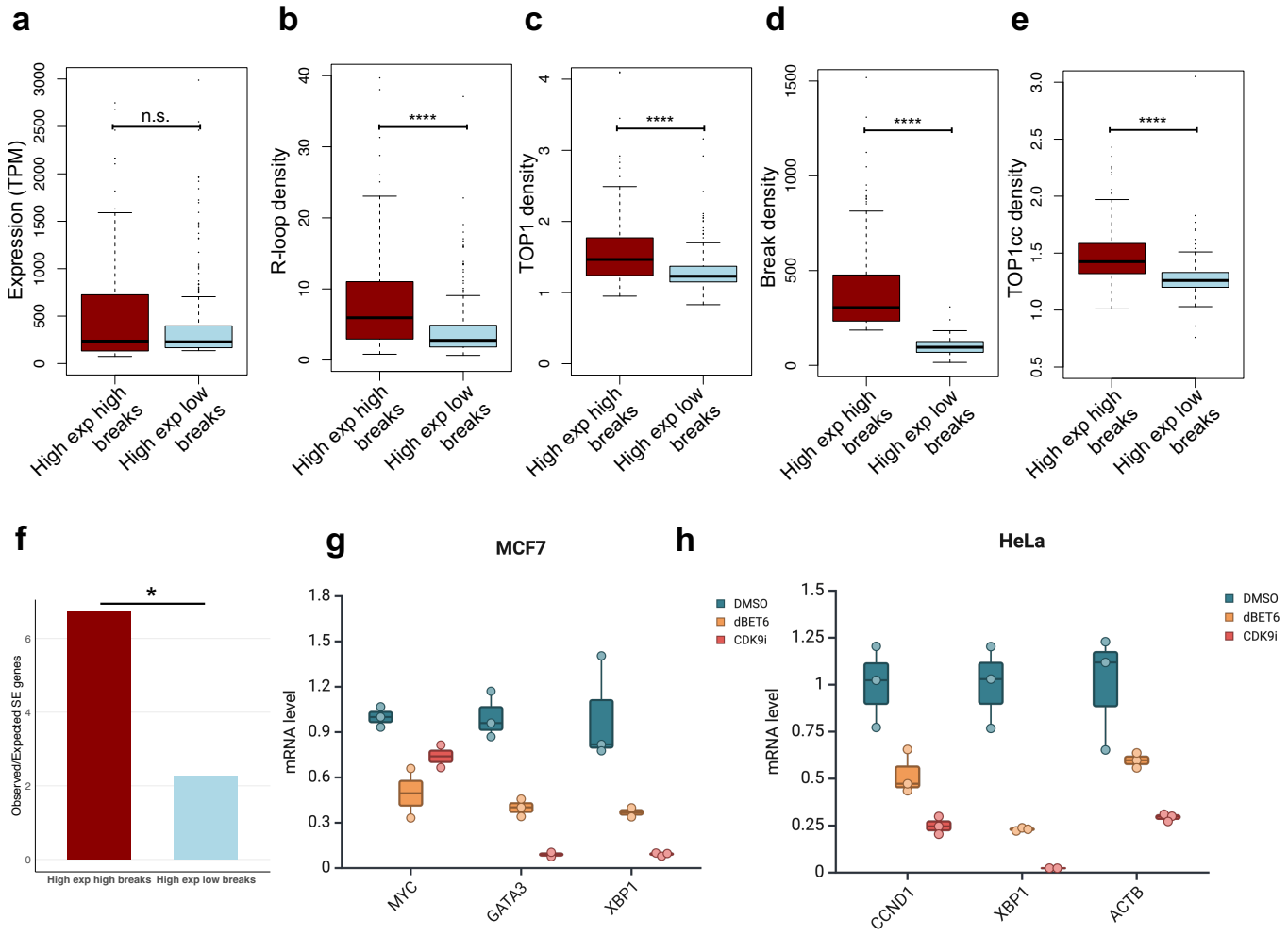
1096 **Extended Data Figure 1. Characterization of transcription stress sites defined by HMM**
1097 **analysis.**
1098 **a,** Distribution of transcription stress sites identified by HMM across genic and intergenic regions.
1099 Genic regions were defined as gene coordinates extended by ±10 kb flanking regions. **b,**
1100 Proportion of intergenic transcription stress sites identified by HMM overlapping key chromatin
1101 states and super-enhancers. **c,** Enrichment analysis showing observed over expected overlap
1102 of intergenic transcription stress sites with chromatin states and super-enhancers. **d,** Enrichment
1103 analysis of transcription stress sites within different gene categories. **e,** percentage of genes
1104 overlapping transcription stress sites as identified by HMM. **f,** Boxplot displaying expression
1105 levels of noted genes. **g,** Logistic regression analysis testing the predictive power of SE status,
1106 gene expression, GC content, and gene length for the presence of transcription stress sites
1107 within genes.


a
Breaks density
Ctrl
RNase H OE
TOP1 KD
Break Density
Row Z-Score
b
Break density
n.s.
n.s.
*
n.s.
Random
non-SE genes
SE genes
G1 rep1
G1 rep2
Cycling rep1
Cycling rep2
c
MCF7 cycling
G1(57.50%)
sub-G1(1.46%)
S(25.30%)
G2-M(15.12%)
Cell count X10
DNA Content (PI Fluorescence Intensity, 561/610 nm)
d
MCF7 G1 synchronized
G1(74.29%)
sub-G1(0.67%)
S(7.94%)
G2-M(16.52%)
Cell count X10
DNA Content (PI Fluorescence Intensity, 561/610 nm)
e
% of BrdU positive cells
****
MCF7 Cycling
MCF7 Serum Starved
f
DAPI
BrdU
Merge
MCF7 Cycling
MCF7 G1
g
all genes
Spearman correlation coefficient: 0.494
P-value: 0
Log10 break density
Log10 Expression
h
Non−SE regulated genes
Spearman correlation coefficient: -0.037
P-value: 3.704e-08
Log10 break density
Log10 Expression
i
SE Regulated Genes
Spearman correlation coefficient: 0.451
P-value: 3.589e-30
Log10 Break Density
Log10 Expression

1108

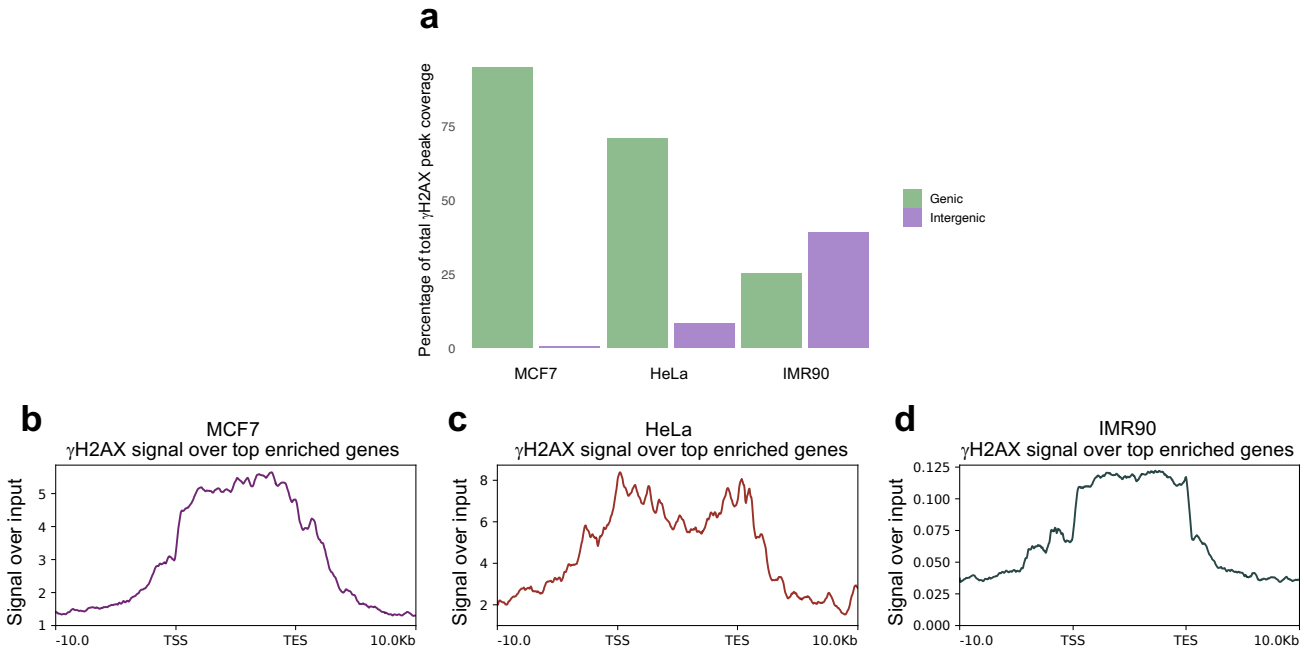
Extended Data Fig. 2

a, Heatmap illustrating break density of SE-regulated genes following RNase H overexpression and TOP1 KD. **b**, Boxplot showing break density of random genes or SE-regulated genes in MCF7 cells synchronized in G1 phase or cycling cells. **c-d,** Flow cytometry analysis of DNA content in MCF7 cells. (c) DNA content profile of asynchronously cycling MCF7 cells. (d) DNA content profile of MCF7 cells synchronized in G1 phase, demonstrating enrichment of the population in G1. Cells were stained with propidium iodide (PI) and analyzed by flow cytometry. **e**, percentage of BrdU positive cells in cycling vs. G1 synchronized MCF7 cells. **f**, representative images of BrdU incorporating cells. **g-l**, Scatter plot demonstrating the correlation between expression and break density for all genes (**g**), non-SE-regulated genes (**h**), and SE-regulated genes (**i**). Boxplot: center line indicates median; box limits represent upper and lower quartiles; whiskers extend to 1.5x interquartile range; dots indicate outliers. P-value was calculated using Mann Whitney U test. n.s. indicating $p > 0.05$, * indicating $p < 0.05$.

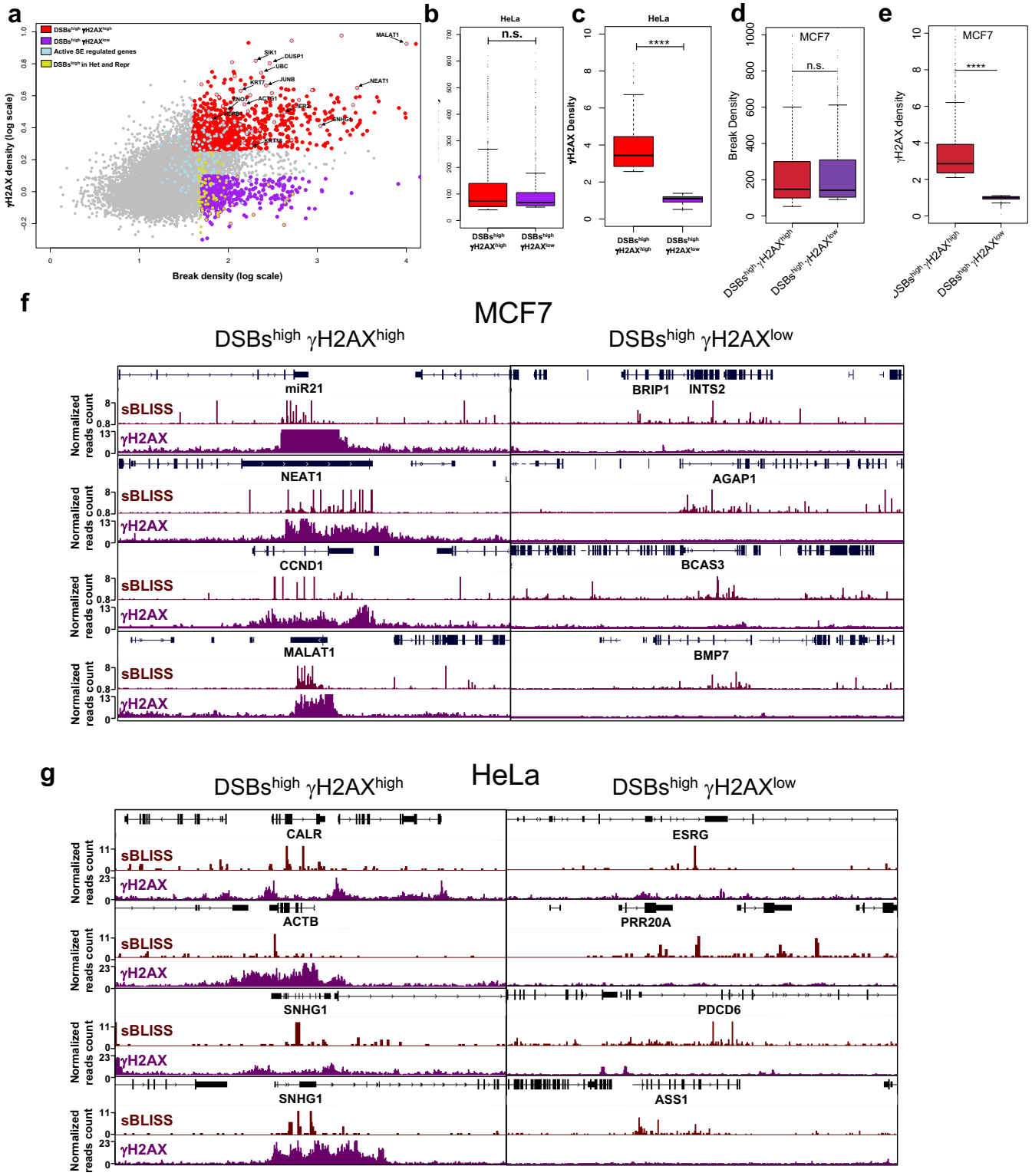


Extended Data Fig. 3

a-e, Boxplot showing expression levels (**a**), R-loop density (**b**), TOP1 density (**c**), break density (**d**), and TOP1cc density (**e**) for high expression high breaks genes versus high expression low breaks genes. **f**, observed to expected ratio of SE genes in high expression high breaks versus high expression low breaks genes. **g-h**, boxplot illustrating mRNA levels of sample SE-regulated genes in MCF7 (**g**) and HeLa (**h**) cells after dBET6 or CDK9i (100nM for 4 hours). Boxplot: center line indicates median; box limits represent upper and lower quartiles; whiskers extend to 1.5x interquartile range; dots indicate outliers. P-value was calculated using Mann Whitney U test. n.s. indicating $p > 0.05$, **** indicating $p < 0.0001$.



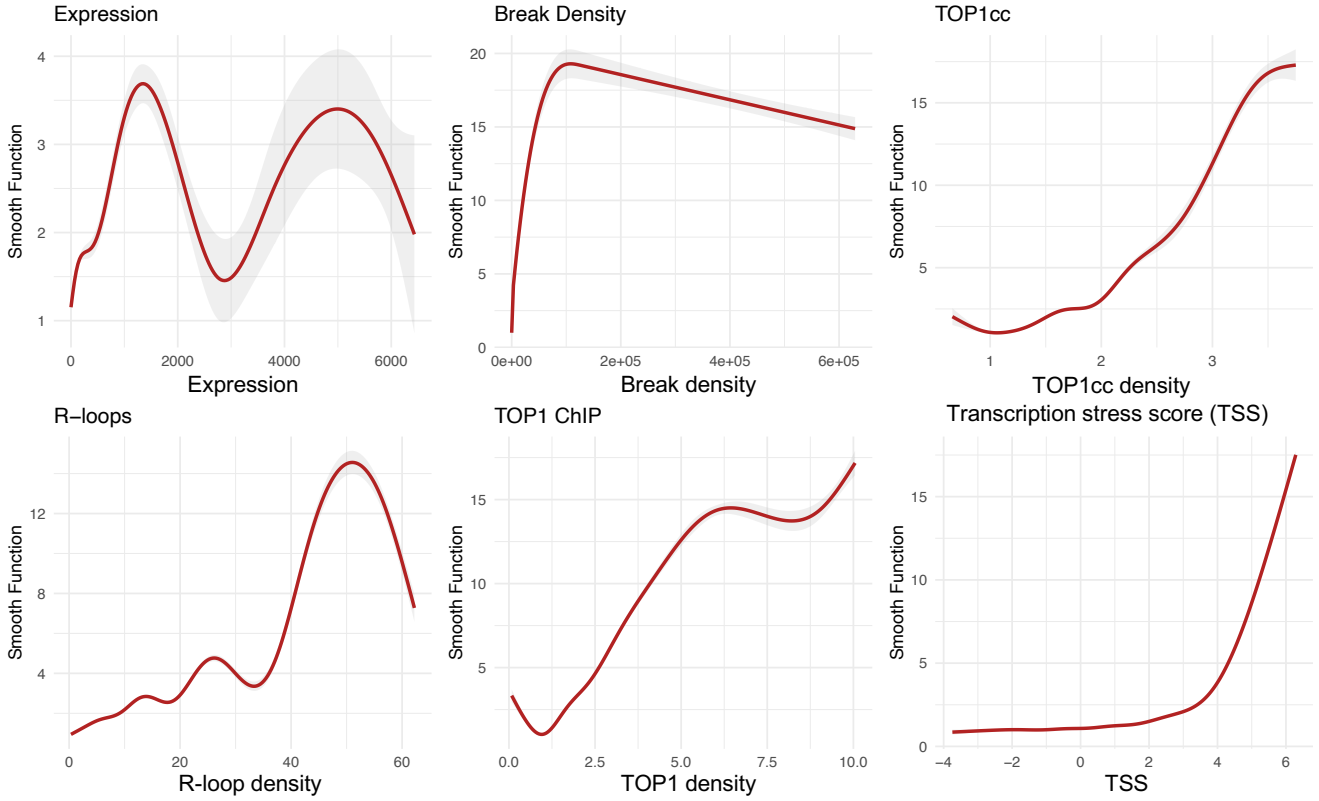
Extended Data Fig. 4.Distribution and enrichment of γH2AX signal at genic regions across different cell lines. a, Percentage of total γH2AX peak coverage occurring within genic and intergenic regions in MCF7, HeLa, and IMR90 cells. Genic regions were defined as gene coordinates extended by ±10 kb flanking regions. (**b**-**d**) γH2AX signal over input across the gene body and flanking regions (±10 kb) for the top enriched genes in (**b**) MCF7, (**c**) HeLa, and (**d**) IMR90 cells. TSS: transcription start site; TES: transcription end site.



Extended Data Fig. 5

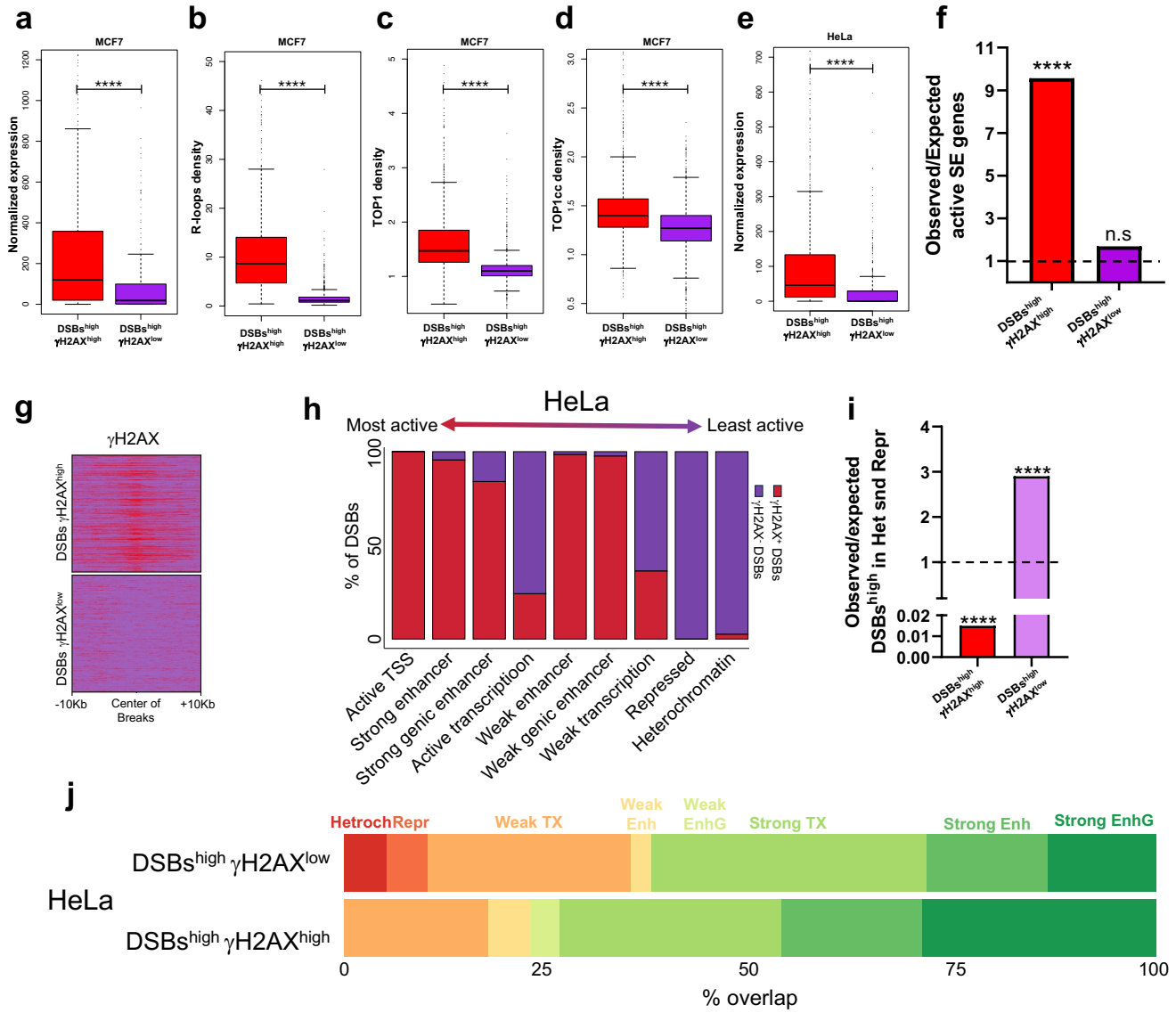
a, Scatter plot showing the correlation between break density and γH2AX density for each gene in HeLa cells. **b-e**, boxplots showing DSBs (b and d) and γH2AX density (c and e) for the two groups of genes defined in (a) and Figure 3F. **d-e,** Genome browser snapshots of representative

examples of DSBs^{high} γH2AX^{high} and DSBs^{high} γH2AX^{low} genes in MCF7 (d), and HeLa (e) cell lines. Boxplot: center line indicates median; box limits represent upper and lower quartiles; whiskers extend to 1.5x interquartile range; dots indicate outliers. P-values were calculated using Mann Whitney U test. n.s. indicating p > 0.05, **** indicating p < 0.0001.



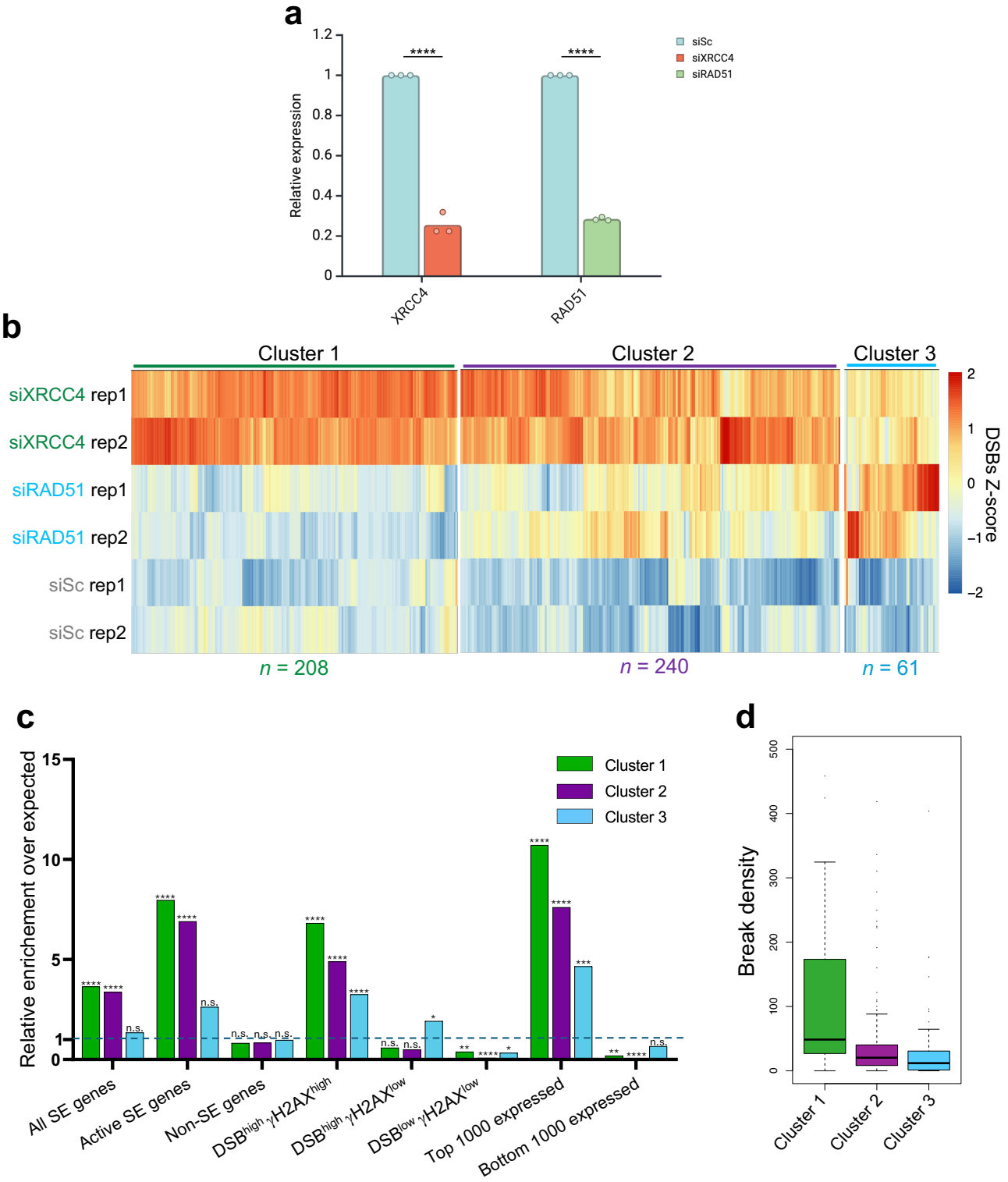
Extended Data Fig.6

Smooth terms from Generalized Additive Models (GAMs) showing the effect of transcription-related predictors on γH2AX density. The y-axis shows the partial effect of each predictor (centered smooth function, s(x)) on γH2AX levels after accounting for non-linear relationships. The red line represents the fitted smooth function, and the shaded gray band indicates the 95% confidence interval.



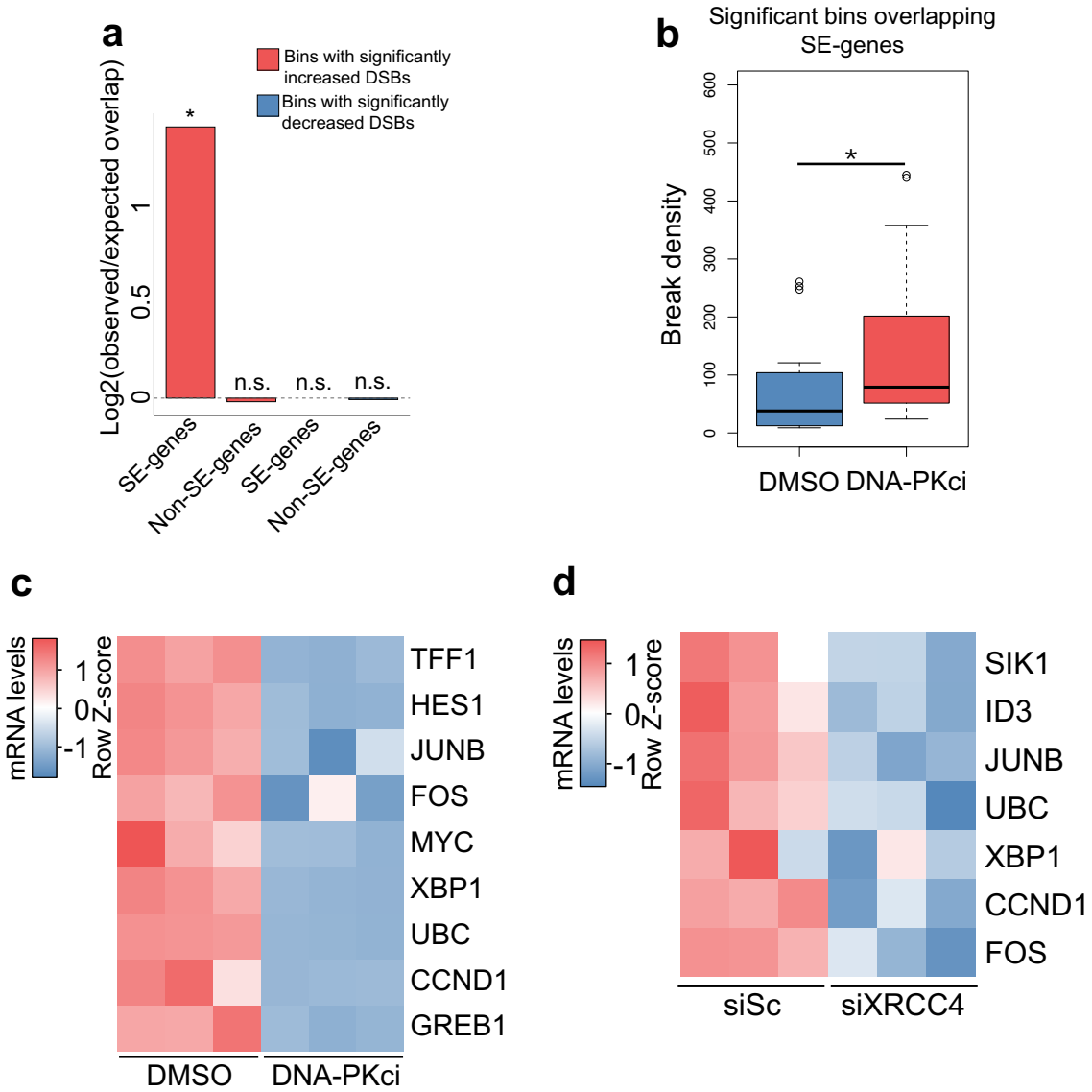
Extended data Fig. 7

a-d, Boxplot showing expression levels (**a**), R-loop density (**b**), TOP1 density (**c**), TOP1cc density (**d**) for DSBs^{high} γH2AX^{high} and DSBs^{high} γH2AX^{low} genes in MCF7 cells. **e**, boxplot showing expression levels for DSBs^{high} γH2AX^{high} and DSBs^{high} γH2AX^{low} genes in HeLa cells. **f**, observed to expected ratio of active SE genes in DSBs^{high} γH2AX^{high} in HeLa cell line. **g**, heatmap showing the enrichment of γH2AX signal at the two defined sets of DSBs. **h**, The percentage of DSBs enriched with γH2AX (γH2AX^{high} DSBs) or lacking γH2AX (γH2AX^{low} DSBs) in each chromatin state of HeLa cell line. Chromatin states were ordered based on what is typically known about their transcriptional activity[73]. **i**, The ratio of observed to expected numbers of high DSB genes overlapping heterochromatin and repressed regions in both gene groups in HeLa cells. **j**, The percentage of overlaps between each gene group and chromatin states in HeLa. For each group, the number of overlaps was counted for each chromatin state and then divided by the total number of overlaps.



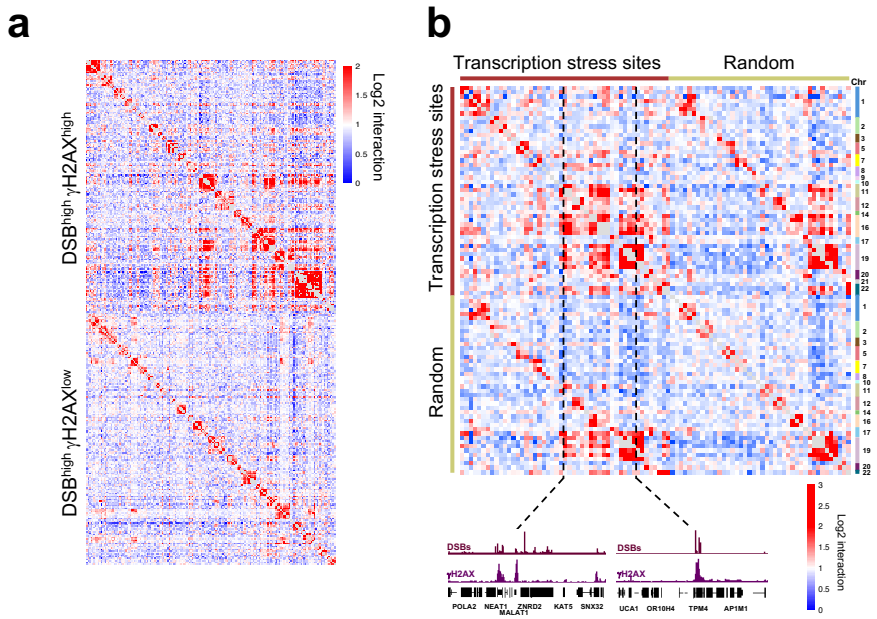
Extended Data Fig. 8.
a, validation of XRCC4 and RAD51 knockdown via RT qPCR, each dot represents the mean of three technical replicates of a biological replicate. b, heatmap showing genes that significantly

accumulated DSBs (p-value <0.05) upon XRCC4 or RAD51 KD. The genes were clustered by hierarchical clustering. c, enrichment of the three gene clusters in different gene sets. d, baseline break density of the three identified clusters.



Extended data Fig. 9: Efficient repair at transcription stress sites is required for SE-regulated oncogene expression

a, Log2 ratio of observed to expected overlap of significantly affected genomic bins (10KB) with SE-regulated genes and non-SE-regulated genes upon DNA-PKci treatment. P-values were calculated using binomial test. **b**, break density of significantly affected bins overlapping SE-regulated genes. **c**, heatmap showing the change of mRNA levels as measured by spike-in qPCR upon DNA-PKci treatment of top γH2AX-enriched SE-regulated genes in MCF7 cells. d, heatmap showing the change of mRNA levels as measured by qPCR upon XRCC4 KD of top γH2AX-enriched SE-regulated genes in HeLa cells. n.s. indicating $p > 0.05$, * indicating $p < 0.05$.



Extended data Fig. 9

a, Heatmap showing Log2 interaction scores for genomic bins containing DSBs^{high} γH2AX^{high} (top) and DSBs^{high} γH2AX^{low}. (bottom). **b**, Heatmap showing Log2 interaction scores for genomic bins containing transcription stress sites or randomly selected bind with similar chromosomal distribution.

Table 1: transcription stress sites genes and whether they have a known oncogenic function.

Gene	Oncogenic function	Source
EGR1	no	
ACTB	yes	Guo C. et al.,[74]
HES1	yes	Wang J. et al.[75]
RNVU1-7	No	
TALAM1	yes	Bermudez G. et al.[76]
MIR21	yes	Zhu S. et al.[77]
RNU4-2	no	
MASCRNA	yes	Xie S. et al.[78]
RNU1-1	no	
HIST2H2AA3	no	
JUNB	yes	Ren F. et al.,[79]
MALAT1	yes	Li Z. et al.[80]
FOS	yes	Ordway J. et al.[81]
ACTG1	yes	Tang Y. et al.[82]
XBP1	yes	OncoKB
HIST2H3A	No	
APRT	No	
HIST1H1C	No	
NEAT1	yes	Yu X. et al.[83]

SNORA10	No	
RPL13A	No	
IER2	yes	Kyjacova L. et al.[84]
MCL1	yes	OncoKB
HIST1H1D	No	
STC2	yes	Qie S. et al.[85]
ID1	yes	OncoKB
SNHG3	yes	Xu b. et al.[86]
CCND1	yes	OncoKB
CDT1	yes	Wang S. et al.[87]
SNORD12B	yes	Zhao X. et al.[88]
HSPB1	Yes	Liang Y. et al.[89]
RPL13	No	
MIR4722	No	
RNU4ATAC	No	
ZFP36	No	
TFF1	yes	Perry J. et al.[90]
IER5	yes	Asano Y. et al.[91]
MIR25	yes	El-Mezayen H. et al.[92]
EGR3	No	
MYC	yes	OncoKB
PTMA	yes	Zhu Y. et al.[93]
CFL1	No	
ELF3	yes	OncoKB
IFITM10	No	
KRT18	No	
MIR3064	no	
SNORA50C	no	
ZNHIT2	No	
TNFRSF12A	yes	Wang C. et.[94]
RNU12	No	
ATF4	yes	Wortel I. et al.[95]
EEF2	yes	Oji Y. et al.[96]
GATA3	yes	OncoKB
TUBA1B	yes	Wang Y. et al.[97]
MLPH	No	
DDX5	Yes	OncoKB
TPM4	No	
RPL32	No	
RPS27	No	
KRT80	yes	Wei X. et al.[98]
NHP2	No	
RPL12	No	
PCBP1	No	
ZFAS1	yes	He A. et al.[99]

SLC7A5	no	

1209