

NR4A1 suppresses cancer replication stress through R-loop-dependent inhibition of immediate early gene transcriptional elongation

Daniel Haber (✉ haber@helix.mgh.harvard.edu)

Massachusetts General Hospital

Hongshan Guo

Massachusetts General Hospital

Gabriel Golczer

Massachusetts General Hospital

Ben Wittner

Massachusetts General Hospital <https://orcid.org/0000-0001-5657-9849>

Adam Langenbacher

Massachusetts General Hospital

Marcus Zachariah

Massachusetts General Hospital

Taronish Dubash

Massachusetts General Hospital Cancer Center

Xin Hong

Massachusetts General Hospital

Valentine Comaills

Massachusetts General Hospital

Risa Burr

Massachusetts General Hospital, Harvard Medical School

Elad Horwitz

Massachusetts General Hospital

Joanna Vuille

Massachusetts General Hospital

Devon Wiley

Massachusetts General Hospital

Brittany Reeves

Massachusetts General Hospital

Jiamin zhang

Massachusetts General Hospital

Kira Niederhoffer

Massachusetts General Hospital

Chenyue Lu

Massachusetts General Hospital Cancer Center and Harvard Medical School

Benjamin Wesley

Massachusetts General Hospital

Uyen Ho

Massachusetts General Hospital

Mehmet Toner

Massachusetts General Hospital, Harvard Medical School, Shriners Hospital for Children

Lee Zou

Massachusetts General Hospital

Raul Mostoslavsky

Massachusetts General Hospital <https://orcid.org/0000-0002-7740-5212>

Shyamala Maheswaran

Massachusetts General Hospital

Michael Lawrence

Massachusetts General Hospital

Biological Sciences - Article

Keywords: NR4A1, cancer replication, R-loop-dependent, inhibition, immediate, early, gene transcriptional elongation

Posted Date: November 2nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-84947/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Deregulation of oncogenic proliferative signals triggers replication stress in cancer cells to which they must adapt^{1,2}. Immediate early genes (IEGs), identified by their rapid stress-induced transient bursts of expression, are critical to integrating downstream signaling pathways^{3,4}. In studying tumor initiation by patient-derived breast cancer cells, we observed acquisition of open chromatin domains at the genebody and 3'-UTR of IEGs, uniquely across the genome. Through *in vivo* and *in vitro* modeling, we show that the IEG and orphan nuclear receptor NR4A1⁵ localizes across multiple IEG genebodies, where it binds to RNA Pol II, arresting transcriptional elongation and generating extensive R-loops and accessible chromatin. Acute stress promptly removes NR4A1 from IEG genebodies, triggering immediate release of their poised transcripts. In breast cancer cells, *NR4A1* overexpression increases tumorigenesis; conversely, its deletion leads to uncompensated replication stress, chromosomal instability and mitotic catastrophe, driven by deregulation of its IEG target *FOS*. A large fraction of primary breast and other cancers exhibit open genebody chromatin at IEGs, consistent with preserved *NR4A1* function. Thus, *NR4A1* mediates a novel transcriptional elongation checkpoint, unique to stress-induced genes and required for their rapid bursts of expression. Cancers that have retained this mechanism in adapting to chronic replication stress may be dependent on *NR4A1* for proliferation.

Main Text

Circulating tumor cells (CTCs) are metastatic precursors that may be viably isolated from blood samples of patients with metastatic breast cancer through microfluidic depletion of normal blood cells^{6,7}. CTCs may be cultured under anchorage-independent conditions and reintroduced into immunosuppressed mice, where they are highly tumorigenic^{6,8}. Patient-derived cultured CTCs demonstrate considerable heterogeneity and plasticity^{6,9}, revealing mechanisms by which they respond to environmental and oncogenic stress, and their associated therapeutic vulnerabilities. By studying early steps in metastatic colonization by CTCs, we uncovered a physiological pathway critical to stress-induced IEG regulation, which is co-opted by cancer cells to suppress replication stress and whose disruption triggers mitotic catastrophe and genomic instability.

Gain of chromatin accessibility at IEG genebodies during tumorigenesis

To study early steps in CTC-induced initiation of metastasis, we tagged two patient-derived, hormone receptor (HR)-positive breast CTC cell lines, BRx142 and BRx82, with both GFP and luciferase and injected them into the left ventricle of immunosuppressed NSG mice, followed by monitoring using *in vivo* luciferase-based imaging (IVIS). An hour after intracardiac inoculation (day 0) and at serial intervals thereafter, mouse tissues were harvested, subjected to single-cell dissociation, and individual tumor cells were collected by GFP-directed fluorescence-activated cell sorting (FACS) (**Fig.1a, Extended Data Fig. 1a, b**). To search for chromatin-associated changes at the earliest possible time points, we applied the Assay

for Transposase-Accessible Chromatin using sequencing (ATAC-seq) to high-purity cancer cell populations (1,000-50,000 cells) from pre-injection cultures and day 0 post-injection tissue harvesting *versus* early metastatic lesions (>day30). Remarkably, we observed few loci with consistent increases in ATAC-seq signal across the genome (20 loci with >3X increase), of which only 5% are at characteristic gene promoters/enhancers indicative of transcriptional activation. Instead, the most striking chromatin accessibility marks acquired during early tumor formation by both CTC cell lines are at genebody and 3'-UTR domains (>82% of loci with increased ATAC-seq) (**Fig. 1b, c**). By gene ontology analysis, the top enriched pathway for these genes is "Cellular response to stimulus" ($P < 8.2E-4$) (**Extended Data Fig. 1c**), and 45% of the genes gaining chromatin accessibility encode known IEGs, including *NR4A1*, *FOS*, *FOSB*, *BTG2*, *BHLHR40*, *MAFF*, *HMGCS1*, *HSPA1A* and *SIK1* (**Fig. 1b**). Among these IEGs, *NR4A1* exhibits the greatest increase in chromatin accessibility (6.5 fold) across its genebody and 3'-UTR (**Fig. 1b**). This unusual IEG genebody open chromatin pattern coinciding with *in vivo* tumor initiation by HR-positive breast CTCs is also evident using the prototypical "triple negative" breast cancer (TNBC) cell line, MDA-MB-231 (hereafter called M231). Intracardiac injection of the highly tumorigenic M231 cells is also accompanied by striking gain of ATAC-seq signal at the genebodies and 3'-UTR of *NR4A1* and other IEGs (**Extended Data Fig. 1d-f**). In contrast to the genebody localization of chromatin accessibility gains during tumor initiation, 18 loci show loss of ATAC-seq signal, but these are localized primarily to intergenic non-coding regions (**Fig. 1b, Extended Data Fig. 1g**).

NR4A1 (also known as *Nur77* or *NGFIB*) has been implicated as a key effector of a wide range of cellular processes, ranging from cancer, metabolism and angiogenesis to inflammation and immune cell differentiation⁵. As an orphan nuclear receptor, *NR4A1* has been thought to mediate these diverse functions as a transcriptional activator, binding to the NGF1-B response elements at gene promoters^{10,11}. The unexpected pattern of *NR4A1* chromatin accessibility led us to test whether its own chromatin binding profile is altered during tumorigenesis. Surprisingly, *NR4A1* ChIP-seq shows minimal (<3%) gene promoter localization in CTC-derived cancer cells (**Fig. 1d**), in marked contrast to other IEGs, like *FOS*, which exhibits 42% localization to promoter regions (**Extended Data Fig. 1h**). Instead, using two different antibodies, we observed massive *NR4A1* ChIP-seq signal across the IEG genebodies, including *NR4A1* itself, overlapping precisely with the ATAC-seq open chromatin domains (**Fig. 1e**). This *NR4A1* localization pattern is not present in pre-injection CTCs or in CTCs recovered from tissues of injected mice at day 0, indicating that it is a phenomenon acquired during early tumor formation. The top enriched pathway for *NR4A1*-bound genes is again "Cellular response to stimulus" ($P < 6.5E-9$), with a predominance of IEGs (**Extended Data Fig. 1i**). Given this unusual chromatin genebody occupancy for a transcription factor, we tested the localization of RNA Polymerase II (RNA Pol II) at IEGs using ChIP-seq. As expected, genome-wide, RNA Pol II is detectable at the transcriptional start sites (TSS) of most genes; however, it accumulates at the genebodies and 3'-UTRs of *NR4A1* and other IEGs, overlapping precisely with *NR4A1* binding itself (**Fig. 1e**). This unusual RNA Pol II distribution is not observed in CTCs prior to tumor formation, suggesting the altered regulation of IEG transcription during tumorigenesis.

As a first step to defining the role of *NR4A1* expression while CTCs transition into early metastatic tumors, we tested the consequences of *NR4A1* suppression and ectopic overexpression on CTC-mediated tumorigenesis. In both BRx142 and BRx82 CTCs, shRNA knockdown of *NR4A1* (67.9% and 68.7% suppression, respectively) dramatically reduces *in vitro* proliferation and *in vivo* tumorigenesis (**Fig. 1f-h, Extended Data Fig. 1j, k**). Conversely, overexpression of *NR4A1* in these CTC lines (40- and 30-fold, respectively) strongly enhances orthotopic tumor formation, as well as intravascular metastasis, while having only a moderate effect on *in vitro* proliferation (**Fig. 1i-k, Extended Data Fig. 1l, m**). *NR4A1* thus appears to regulate a rate-limiting function in both primary and metastatic tumorigenesis.

NR4A1 restrains RNA Pol II transcriptional elongation at IEG genebodies

To better understand the functional role of *NR4A1* and its chromatin configuration, we sought to recapitulate the tumor-associated phenotype using *in vitro* models of cellular stress responses. Among various stimuli (DNA damage, heat shock, serum stimulation), we found serum starvation followed by replenishment, a classical IEG stimulation protocol that recapitulates proliferative and replication stress signals, to elicit *NR4A1* induction and altered localization analogous to that observed *in vivo*. Since CTCs are already maintained in the absence of serum, these experiments were performed in M231 cells, as well as in MCF10A breast epithelial cells, both of which show serum-dependent proliferation. In these cells, sharp peaks of IEG expression are observed between 30-60 min after serum replenishment, with *NR4A1* showing the greatest increase, demonstrated by RNA (>400-fold and >800-fold in MCF10A and M231 cells, respectively) and protein quantitation, and by immunofluorescence staining (**Fig. 2a, Extended Data Fig. 2a-c**). The IEG *FOS* is rapidly induced upon serum starvation, while *FOSB* and *EGR1* peak following serum replenishment (**Fig. 2a**).

By ChIP-seq analysis of cells under baseline serum-replete conditions, IEG gene products such as *FOS* and *MYC*, and the *MYC* heterodimerization partner *MAX* are, as expected, predominately localized to transcriptional start sites (TSS). However, no significant promoter binding by *NR4A1* is observed across the genome. Instead, 78% of chromatin-associated *NR4A1* protein resides at genebody and 3'-UTR regions (**Fig. 2b, Extended Data Fig. 2d, e**). As in the tumor analyses, 14 of 39 *NR4A1*-targeted genebodies encode known IEGs, and the top hit in gene ontology analysis is again "Cellular response to stimulus" ($P < 4.0E-10$) (**Extended Data Fig. 2f, Supplementary Table 1**). Serial *NR4A1* ChIP assays in both MCF10A and M231 cells cultured under baseline conditions, serum starvation, and at closely spaced time points following serum replenishment show striking temporal dynamics in *NR4A1* localization across IEGs: *NR4A1* binding to the genebodies and 3'-UTR is pronounced at baseline, modestly reduced upon serum starvation, and then disappears within 30 min of serum replenishment, beginning to reappear at 6 hrs (**Fig. 2c, Extended Data Fig. 2g, h**). H3K27ac ChIP-seq at these time points reveals no associated

enhancer activity in NR4A1-bound regions, nor do the promoters of NR4A1-targeted genes show changes in H3K27ac marks, indicating that NR4A1 is not acting as a prototypical transcriptional activator at promoter or enhancer regions (**Fig. 2c, Extended Data Fig. 2g**).

To correlate the sequential pattern of NR4A1 loading at IEG genebodies with that of RNA polymerase, we also performed RNA Pol II ChIP-seq at serial intervals following serum starvation and refeeding. Across the genome, in both MCF10A and M231 cells, RNA Pol II ChIP-seq reads are strongly enriched at TSS under all culture conditions (**Extended Data Fig. 2i**). In stark contrast, across IEGs under baseline culture conditions, RNA Pol II residency is evident broadly across genebodies and at 3'-UTRs, precisely overlapping with the presence of NR4A1, consistent with our observations in CTC-derived tumors (**Fig. 2c, Extended Data Fig. 2g**). The localization of RNA Pol II along the IEG genebodies is resolved 30-60 min after serum replenishment, coinciding with disappearance of NR4A1 occupancy. Remarkably, both NR4A1 and RNA Pol II binding to genebodies are inversely correlated with the RNA expression of IEGs: IEG expression is virtually undetectable when NR4A1 and RNA Pol II are bound at IEG genebodies, and it peaks 30 min after their release from these sites (60 min after serum addition) (**Fig. 2d**). This tight and inverse sequential timing suggests an inhibitory role for NR4A1 on IEG transcription, associated with RNA polymerase stalling in these genebody regions. The genebody ChIP-seq patterns are observed using antibodies against total RNA Pol II, as well as against the phosphorylated forms of the polymerase C-Terminal Domain (CTD) (phospho-S2 and phospho-S5). Specific RNA Pol II phosphorylation sites have been best studied in the context of MYC-associated promoter-proximal pausing, with S5-phosphorylation characteristic of the paused polymerase and S2-phosphorylation observed with the licensed and transcriptionally active enzyme¹². The presence of polymerase phosphorylated at both residues is unexpected, and it raises the possibility that the licensed enzyme is further modified at stalling sites along the genebody. To further confirm the RNA polymerase ChIP-seq pattern, we used a different set of antibodies against total RNA Pol II or RNA Pol II phospho S2 to perform ChIP-qPCR. Again, abundant signal of both total RNA Pol II and RNA Pol II phospho S2 is evident at IEG genebodies under basal culture conditions, with a marked reduction 1 hr after serum replenishment (**Extended Data Fig. 2j**).

The virtually immediate induction of IEG expression following acute stress stimuli is matched by its very rapid resolution (i.e. a single burst of expression). Interestingly, RNA Pol II does not reload at either the IEG TSS or genebodies before 6 hrs after serum stimulation (**Fig. 2c, Extended Data Fig. 2g, j**). This suggests that IEG expression is primarily mediated by the rapid completion of stalled transcription, which results in a burst of gene expression triggered by acute stress. This "pause and release" pattern may therefore account for both the precipitous onset and rapid termination of IEG expression. Supporting this model, after serum deprivation and replenishment, precision nuclear run-on followed by sequencing (PRO-seq)¹³, shows a transient increase in nascent transcription across IEG genebodies, coinciding with the reduction in RNA Pol II localization (**Fig. 2e**). Finally, the RNA polymerase travelling ratio (TR), calculated by

comparing RNA Pol II ChIP-seq read density between promoter and genebody regions at different time points¹⁴ (see Methods), demonstrates an anti-correlation between NR4A1 binding intensity and RNA Pol II TR, again consistent with a functional role for NR4A1 in inhibiting transcriptional elongation of IEGs (**Fig. 2f**).

While NR4A1 can bind to DNA with sequence specificity, ChIP-seq analyses do not identify a specific consensus sequence for its localization along the IEG genebodies. We therefore postulated that the recruitment of NR4A1 to these genebodies may result from protein-protein interactions, potentially including RNA Pol II itself. To test for such a physical interaction, we immunoprecipitated NR4A1 from MCF10A cells, followed by western blotting against the activated phospho-S2 residue within the RNA Pol II CTD. Under basal serum conditions, phospho-S2 RNA Pol II co-immunoprecipitates with NR4A1, but this protein association is no longer detectable upon serum deprivation, as NR4A1 starts to dissociate from the IEG genebodies (**Fig. 2g**). Thus, NR4A1 dynamically interacts with the RNA polymerase elongation complex, inhibiting its activity at baseline and releasing its inhibition in response to stress-induced signals. A similar protein association *in vivo* between NR4A1 and RNA Pol II is evident in CTC-derived tumors (**Fig. 2h**).

NR4A1-dependent R-loops contribute to chromatin accessibility and transcriptional arrest at IEG genebodies

To determine whether NR4A1 directly mediates RNA Pol II stalling at IEG genebodies, we used CRISPR/Cas9 with different pairs of guide-RNAs to generate multiple *NR4A1* knockout clones in M231 cells (see Methods). Genomic PCR demonstrated successful deletion of the targeted *NR4A1* gene fragments, and the absence of NR4A1 protein was confirmed by western blotting (**Extended Data Fig. 3a, b**). ChIP-seq analysis of *NR4A1*-null M231 cells under baseline culture conditions shows no change in overall RNA Pol II presence and localization across the genome, but a striking reduction of the polymerase at the genebody regions of IEGs (**Fig. 3a**). Consistent with the role of NR4A1 in restraining transcriptional processing of IEGs, expression of *FOS*, *FOSB* and *EGR1* is markedly elevated in *NR4A1*-null cells cultured under baseline conditions, as well as following serum withdrawal and replenishment (**Fig. 3b, Extended Data Fig. 3c**). These findings were verified using PRO-seq, showing increased transcription across genebody regions of IEGs, correlated with the reduction in pile-up of RNA Pol II, in *NR4A1*-null M231 cells, compared with parental cells, under baseline culture conditions (**Fig. 3c, d, Extended Data Fig. 3d, e and Supplementary Table 2**). Unlike M231 cells, MCF10A cells do not tolerate CRISPR-mediated deletion of *NR4A1*. We therefore achieved shRNA-mediated *NR4A1* knockdown in these cells using two different sequences (77.7% and 91.4% knockdown, respectively) (**Extended Data Fig. 3f**). Again, RNA Pol II ChIP-seq following NR4A1 knockdown shows a striking reduction in RNA Pol II localization to IEG

genebodies under baseline culture conditions, associated with increased expression of IEGs, including *FOS* and *EGR1* (**Extended Data Fig. 3g, h**).

Stalled RNA polymerase processing may be associated with hybridization of the nascent transcript with the unwound matching antisense DNA strand, forming DNA-RNA hybrid structures called R-loops¹⁵. Such structures may result from RNA Pol II pausing, but they could also directly contribute to pausing^{16,17}. To determine whether *NR4A1*-mediated RNA Pol II pausing is associated with R-loop formation, we first undertook DNA-RNA immunoprecipitation (DRIP) using the canonical S9.6 monoclonal antibody, which recognizes DNA-RNA hybrids with subnanomolar affinity^{18,19}. Using DRIP followed by sequencing (DRIP-seq) or qPCR (DRIP-qPCR), we observed strong R-loop signals across the genebodies of IEGs, including *FOS* and *EGR1*, in both M231 and MCF10A cells, under baseline and serum starved conditions, with virtual disappearance of the R-loop signal upon serum stimulation (**Fig. 3e, Extended Data Fig. 4a**). Similarly, in the *in vivo* CTC-derived tumors characterized by IEG genebody chromatin accessibility and by *NR4A1* and RNA Pol II co-localization, S9.6 DRIP-qPCR analysis reveals dramatic enrichment of R-loop signal at the same IEG genebody loci (**Fig. 3f, Extended Data Fig. 4b**). In *NR4A1*-null M231 cells under baseline conditions, DRIP-seq reveals a dramatic erasure of IEG genebody R-loops (**Fig. 3g**), and a comparable reduction of R-loops at IEG genebodies is observed in MCF10A cells following shRNA-mediated *NR4A1* knockdown (**Extended Data Fig. 4c**). In all these experiments, specificity of the R-loop signal was confirmed using *in vitro* RNase H digestion. Thus, *NR4A1* localization to the IEG genebodies drives the accumulation of R-loops at these loci. The reduction in R-loops in *NR4A1*-null M231 cells is associated with increased expression of IEGs, notably the prototypical IEG *FOS* (**Fig. 3b, Extended Data Fig. 3d**). Since R-loop formation may be either cause or consequence of delayed transcriptional processing²⁰, we tested the direction of this causation by establishing ectopic expression in M231 cells of RNase-H1 (**Extended Data Fig. 4d**), which directly degrades the RNA strand in the DNA-RNA hybrids and resolves R-loops *in vivo*^{19,21,22}. S9.6 DRIP-qPCR analysis of RNase-H1-expressing cells growing under baseline culture conditions shows abrogation of the R-loop signal at the *FOS* (>13-fold reduction) and *EGR1* (>5-fold reduction) genebodies (**Extended Data Fig. 4e**). *In vivo* RNase-H1 expression also leads to significantly increased expression of *FOS*, and other IEGs (**Fig. 3h, Extended Data Fig. 4f**), indicating that R-loops contribute to transcriptional suppression.

Abundant R-loops along the genebody of IEGs have the potential to disrupt DNA compaction, resulting in more accessible chromatin. We therefore asked if the striking genebody ATAC-seq signal that initially drew our attention to *NR4A1* and IEGs could itself be the result of extensive R-loop formation. MCF10A cells under baseline culture conditions show ATAC-seq signal across the IEG genebodies, comparable to that observed in CTC-derived tumors (**Fig. 3i**). *In vivo* expression of RNase H1 in these cells leads to a marked diminution in this genebody chromatin accessibility (**Fig. 3i, Extended Data Fig. 4g**), indicating that it is indeed a consequence of R-loop accumulation. *NR4A1* suppression in MCF10A cells results both

in the resolution of IEG genebody R-loops as well as reduced ATAC-seq signal at these loci, supporting the role of NR4A1 binding in mediating these two phenomena (**Fig. 3j, Extended Data Fig. 4c**). Direct resolution of R-loops through *in vivo* expression of RNase-H1 also reduces RNA Pol II occupancy at the *FOS* genebody in MCF10A cells (**Extended Data Fig. 4h**), while NR4A1 binding itself is not affected (**Extended Data Fig. 4i**). The direct effect of R-loops on IEG gene expression is consistent with the observation that the small-molecule transcriptional inhibitor 5,6-dichloro-1- β -D-ribofuranosylbenzimidazole (DRB) induces comparable RNA Pol II pausing and R-loop signal along the IEG genebodies, irrespective of *NR4A1* gene status in M231 cells (**Extended Data Fig. 4j, k**).

Taking all of this evidence together, we propose a model whereby NR4A1 binds to RNA Pol II, arresting transcriptional elongation specifically along IEG genebodies, generating extensive R-loops that result in dramatic genebody chromatin accessibility. This phenomenon leads to a piling up of poised IEG transcripts. When NR4A1 dissociates from IEG genebodies and releases RNA Pol II in response to acute stress signals, these poised transcripts are very rapidly completed and released. This unique IEG transcriptional elongation checkpoint thus triggers an immediate, coordinated and transient burst of stress-induced IEG expression (**Fig. 3k**).

NR4A1 suppresses mitotic catastrophe

The mechanism whereby *NR4A1* controls the acute IEG stress response may also underlie its potent role in CTC-mediated tumorigenesis (**Fig. 1f-k**). Much as cultured CTCs do not tolerate loss of *NR4A1* expression, *NR4A1*-null M231 cells show impaired proliferation *in vitro* and retarded tumorigenesis *in vivo* (**Extended Data Fig. 5a-d**). MCF10A cells also show significantly reduced proliferation *in vitro* following *NR4A1* knockdown (**Extended Data Fig. 3f, 5e**). RNA-seq in *NR4A1*-null M231 cells compared with parental cells identifies 501 downregulated genes (fold change >2, $P < 0.05$), with enrichment for pathways involved in DNA replication ($P < 3.0E-5$) and DNA repair ($P < 5.7E-4$) (**Extended Data Fig. 5f, g, Supplementary Table 3**). Consistent with these findings, *NR4A1*-null cells show elevated levels of phosphorylated Chk1, γ -H2AX and RPA32, evidence of replication stress and activated DNA damage responses (**Fig. 4a, Extended Data Fig. 5h, i**). The increased Chk1 levels in *NR4A1*-null cells are associated with increased sensitivity to the Chk1 inhibitor, MK-8776, compared with control cells (**Fig. 4b**).

Underlying this activation of DNA damage response pathways are massive mitotic defects in *NR4A1*-deficient cells. Dual staining for α -tubulin and DAPI in *NR4A1*-null M231 cells reveals large numbers of chromosomal defects: a mean 39.5% of cells (range: 34.1%-44.9%) have multiple nuclei or gross chromosomal fragmentation, compared to parental cells with a mean 2.1% (range: 1.1%-3.0%, $P = 0.0028$) (**Fig. 4c, d**). Similar chromosomal instability, including cells with multiple nuclei and micronuclei, is

observed in MCF10A cells upon shRNA knockdown of NR4A1 (**Fig. 4e, f**). Flow cytometric analysis of DNA content shows a very high degree of aneuploidy and genome duplications in *NR4A1*-deficient M231 cells, compared with parental cells (**Fig. 4g**). Single-cell karyotypes show increases in the number of chromosomes per cell from a median 59 in parental M231 clones to 87, 132 and 182 in three independent *NR4A1*-null clones, respectively ($P=0.015$, $P<0.0001$ and $P<0.0001$, respectively) (**Fig. 4h**). *NR4A1*-null M231 cells also show a prolonged G2/M phase consistent with mitotic delay, both under baseline culture conditions and following serum replenishment of starved cells, as measured by 5-ethynyl-2'-deoxyuridine (EdU) incorporation assay (**Extended Data Fig. 6a**). In addition to chromosomal defects evident in stably generated *NR4A1*-null cells, acute *NR4A1* knockdown using three different shRNA constructs in M231 cells (68.8%, 72.8% and 96.0% knockdown, respectively) also triggers genomic duplications and G2/M arrest (**Extended Data Fig. 6b-e**). Similar chromosomal defects are observed following knockdown of *NR4A1* in CTC cell lines (**Extended Data Fig. 6f**). Together, these findings indicate that *NR4A1* expression is critical to maintaining genomic stability and that its absence results in major chromosomal defects that compromise cell proliferation and survival.

To test whether the genomic instability resulting from *NR4A1* deletion results in part from its role in controlling stress-induced IEG expression, we tested the consequences of individual IEG knockdown in *NR4A1*-null M231 cells. Remarkably, knockdown of *FOS* using either siRNA or shRNA in *NR4A1*-depleted cells effectively reverses their massive mitotic defects, genome duplication and proliferative failure (**Fig. 4i, Extended Data Fig. 7a-c**). No such effect is observed following knockdown of other IEGs tested, including *FOSB*, *BHLHE40* and *MYC* (**Fig. 4i**). The reversal of genome duplication in *NR4A1*-null cells is most likely due to the death of already mitotically compromised cells, and the suppression of further genomic instability mediated by deregulated *FOS* expression. Indeed, suppressing *FOS* in *NR4A1*-null M231 cells leads to reduced replication stress and enhanced proliferation of cells with largely corrected chromosomal content (**Fig. 4j, k**). Consistent with these findings, ectopic expression of *FOS* in wild-type M231 cells suffices to trigger replication stress and mitotic defects (**Extended Data Fig. 7d-i**). Indeed, *FOS* together with other IEGs are significantly induced when *NR4A1* is suppressed using shRNA in CTCs; conversely, their expression is suppressed when *NR4A1* is overexpressed in CTC-derived tumors, which exhibit reduced DNA damage and apoptosis, compared with parental CTC-derived tumors (**Extended Data Fig. 8a-f**). Thus, restriction of inappropriate *FOS* expression, above all other IEGs, appears to be linked to the function of *NR4A1* in suppressing replication stress and chromosomal instability.

Prevalence of NR4A1 and IEG genebody chromatin accessibility in primary human cancers

Having described the phenomenon of IEG genebody and 3'-UTR chromatin accessibility acquired by cultured breast cancer cells during *in vivo* tumorigenesis, we sought to determine whether this phenomenon is observed in primary human cancers. By reanalyzing a recent ATAC-seq analysis of 404

TCGA human primary cancers²³, we find *NR4A1* and IEG genebody chromatin accessibility to be prevalent across different cancer subtypes, ranging from 100% of prostate cancers to 59.5% of breast cancers and 11.8% of liver cancers (genebody/promoter ATAC-seq ratio >1; **Fig. 5a, b, Extended Data Fig. 9a**). Across all TCGA cancers studied, *NR4A1* genebody ATAC-seq signal shows strong concordance with ATAC-seq signal at other IEG genebodies, including *FOS* and *EGR1* (**Fig. 5c, Extended Data Fig. 9a**). Among primary breast cancers, *NR4A1* genebody ATAC-seq signal is highly detectable in 65.9% of luminal A/B cancers, which are characteristically HR-positive and well differentiated, versus 22.7% of basal cancers, which include the more aggressive TNBC, and 9.1% of HER2-amplified subtypes (**Fig. 5d**). Reanalyzing a published dataset of ATAC-seq from normal mouse mammary gland development²⁴, we also find IEG genebody accessibility in normal mouse mammary cells (**Extended Data Fig. 9b**). Thus, this IEG chromatin pattern appears to reflect normal physiological mechanisms of IEG regulation, which are preserved in more differentiated breast cancers that have retained replication stress control pathways. Consistent with this concept, the presence of *NR4A1* genebody chromatin accessibility is associated with lower stages of breast cancer, and it is highly correlated with a favorable clinical outcome (P=0.013) (**Fig. 5e, f**). Across all cancer types, *NR4A1* genebody open chromatin is correlated with reduced apoptosis signaling, stress response and DNA damage response signaling, with increased developmental growth signaling pathways including TGF- β , WNT and NOTCH, as well as with estradiol response signaling, predominately in breast cancers (**Fig. 5g, Extended data Fig. 9c**). Of note, the *NR4A1*-associated open chromatin domains, correlated with reduced IEG expression, are functionally distinct from overexpression of total *NR4A1* mRNA, which is generally associated with an adverse clinical outcome across different cancer types²⁵⁻²⁷.

Finally, to test the potential therapeutic implications of NR4A1 targeting in breast cancer, we treated NSG mice with CTC-derived orthotopic mammary tumors with the NR4A1 small-molecule inhibitor DIM-C-pPhCO₂Me (NR4A1-i). NR4A1 inhibition dramatically suppresses tumorigenesis by these patient-derived breast cancer cells (**Fig. 5h**). The residual tumors in NR4A1-i-treated mice show elevated expression of the DNA damage marker phospho γ H2AX, and decreased expression of the proliferation marker Ki-67 (**Fig. 5i, j, Extended Data Fig. 9d**).

Perspectives

We have uncovered a new transcriptional elongation checkpoint, mediated by *NR4A1*, which is specific to IEGs, whose exceptionally rapid bursts of expression depend on the immediate release of poised transcripts (**Fig. 5k**). This non-canonical mechanism of transcriptional regulation by an orphan nuclear receptor has implications both for the physiological response to stress in normal cells, as well as its adaptation in cancer. The concept of non-oncogene addiction describes the dependence of cancer cells on genes that are not themselves drivers of proliferation, but which control regulatory pathways critical to

cancer cell survival. The dramatic effect on tumorigenesis of *NR4A1* overexpression and knockdown suggests that cancer cells may rely on this pathway to mitigate replication stress resulting from aberrant proliferative signals. This finding raises the possibility of therapeutic targeting of *NR4A1*, potentially in combination with inducers of cellular stress, particularly in the >60% of cancers that appear to show preservation of the *NR4A1*-dependent transcriptional elongation checkpoint. While analysis of TCGA data indicates that more differentiated cancer types are more likely to have preserved the IEG regulatory signatures reported here, we note the effectiveness of *NR4A1* suppression in cultured CTCs from advanced HR-positive breast cancer and in the highly malignant M231 TNBC cells, which suggests that this pathway may also be relevant in advanced breast cancers.

We also note that *NR4A1* has been implicated as a key regulator of T cell exhaustion, the unresponsive phenotype that follows excessive stimulation in antigen-reactive T cells, as well as in synthetic CAR-T cells^{28,29}. IEG induction accompanies T cell activation, and *NR4A1* has been postulated to block FOS-JUN promoter binding sites and suppress AP-1-mediated transcription. However, our review of *NR4A1* binding landscapes in these cells²⁹, suggests predominant binding to the genebodies, rather than to the promoters of IEGs, consistent with the transcriptional elongation control described here. Most recently, *NR4A1* was also reported to be one of the key factors to restrain B cell responses to antigen³⁰. Taken all together, these observations point to potentially convergent mechanisms in immune and cancer cells, with *NR4A1*-mediated adaptation to chronic antigen stimulation leading to an exhaustion phenotype in immune cells, and *NR4A1*-mediated tolerance to oncogene-driven replication stress preventing mitotic catastrophe in cancer cells. The dependence of critical IEG-mediated signals on *NR4A1* may thus reveal therapeutic opportunities in both drug-based and immune cell-mediated treatments of cancer.

The genebody- and 3'-UTR-centered transcriptional elongation arrest mediated by *NR4A1* differs fundamentally from the more general function of *MYC* in releasing the common pausing of RNA polymerase 30-50 bp downstream of the TSS^{14,31,32}. *MYC*-regulated promoter-proximal pausing release serves both to prevent leaky transcription and as a rheostat to broadly increase transcription under proliferative conditions. In contrast, the more targeted effect of *NR4A1* across the genebodies and 3'-UTRs of IEGs enables their rapid and synchronous expression in response to stimuli. This immediate and limited burst of coordinated IEG expression is essential to multiple stress responses. Recently, transcription inhibitors have been shown to trigger R-loop formation along the genebody of highly expressed genes²⁰, a finding that is consistent with our observation of these DNA-RNA hybrids across the genebodies of IEGs whose transcription is blocked by *NR4A1* binding. R-loops have been primarily studied as consequences of DNA damage-induced collisions between transcriptional and replication machineries¹⁵, attempted transcription through heterochromatin barriers³³, or abnormalities in cleavage and polyadenylation (CPA) factors³⁴. In contrast to these pathology-associated R-loops, those induced by *NR4A1* localization to IEG genebodies appear to be linked to the normal physiological regulation of IEG

gene expression, and their removal through RNase-H1 treatment leads to aberrant IEG expression. Notably, the generation of 3'-UTR-localized R-loops has been shown to recruit repressive histone modifiers capable of suppressing transcriptional termination³³. Together with our description of R-loop-dependent IEG regulation in normal and cancer cells, these findings raise the possibility of site-specific manipulation of transcriptional processing.

Declarations

Author contributions

H.G., S.M., M.S.L. and D.A.H. conceived the project, provided leadership for the project and drafted the manuscript. H.G., M.Z., X.H., V.C., T.D.D., E.H., R.B., J.A.V., D.F.W., B.A.R., J.Z., K.L.N., B.W. and U.H. conducted all the experiments. H.G., G.G., B.S.W., A.L., C.L. and M.S.L. carried out all the data analysis. M.T. developed CTC-iChip technology. L.Z. and R.M. contributed expertise in interpreting the results. All authors edited the manuscript.

Competing interests

Massachusetts General Hospital (MGH) has applied for patents regarding the CTC-iChip technology and CTC detection signatures. M.T., D.A.H. and S.M. are cofounders and have equity in Tell-Bio, which is not related to this work. The interests of these authors were reviewed and managed by MGH and Partners HealthCare in accordance with their conflict of interest policies. All other authors declare no competing interests.

Acknowledgments

We thank L. Libby for technical support; J. Fung for flow sorting assistance and M. Guo for graphic design. We thank N. Dyson and members in Haber/Maheswaran lab for discussions. This work was supported by grants from National Institute of Health (2R01CA129933 to D.A.H, 2U01EB012493 to M.T., D.A.H., S.M., 5U01EB012493 to M.T., 5P41EB002503 to M.T., R01CA197779 to L.Z. and R01GM128448 to R.M.), Howard Hughes Medical Institute (to D.A.H.), ESSCO Breast Cancer Research Fund (to S.M.), Breast Cancer Research Foundation (to D.A.H.) and National Foundation for Cancer Research (to D.A.H.).

References

1. Kotsantis, P., Petermann, E. & Boulton, S. J. Mechanisms of Oncogene-Induced Replication Stress: Jigsaw Falling into Place. *Cancer Discov* **8**, 537–555, doi:10.1158/2159-8290.CD-17-1461 (2018).

2. Macheret, M. & Halazonetis, T. D. DNA replication stress as a hallmark of cancer. *Annu Rev Pathol* **10**, 425–448, doi:10.1146/annurev-pathol-012414-040424 (2015).
3. Morgan, J. I. & Curran, T. Immediate-early genes: ten years on. *Trends Neurosci* **18**, 66–67 (1995).
4. Bahrami, S. & Drablos, F. Gene regulation in the immediate-early response process. *Adv Biol Regul* **62**, 37–49, doi:10.1016/j.jbior.2016.05.001 (2016).
5. Maxwell, M. A. & Muscat, G. E. The NR4A subgroup: immediate early response genes with pleiotropic physiological roles. *Nucl Recept Signal* **4**, e002, doi:10.1621/nrs.04002 (2006).
6. Yu, M. *et al.* Ex vivo culture of circulating breast tumor cells for individualized testing of drug susceptibility. *Science* **345**, 216–220, doi:10.1126/science.1253533 (2014).
7. Maheswaran, S. & Haber, D. A. Ex Vivo Culture of CTCs: An Emerging Resource to Guide Cancer Therapy. *Cancer Research* **75**, 2411–2415, doi:10.1158/0008-5472.can-15-0145 (2015).
8. Ozkumur, E. *et al.* Inertial focusing for tumor antigen-dependent and -independent sorting of rare circulating tumor cells. *Sci Transl Med* **5**, 179ra147, doi:10.1126/scitranslmed.3005616 (2013).
9. Jordan, N. V. *et al.* HER2 expression identifies dynamic functional states within circulating breast cancer cells. *Nature* **537**, 102–106, doi:10.1038/nature19328 (2016).
10. Pei, L. *et al.* NR4A orphan nuclear receptors are transcriptional regulators of hepatic glucose metabolism. *Nat Med* **12**, 1048–1055, doi:10.1038/nm1471 (2006).
11. Winoto, A. & Littman, D. R. Nuclear hormone receptors in T lymphocytes. *Cell* **109 Suppl**, S57–66, doi:10.1016/s0092-8674(02)00710-9 (2002).
12. Phatnani, H. P. & Greenleaf, A. L. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev* **20**, 2922–2936, doi:10.1101/gad.1477006 (2006).
13. Mahat, D. B. *et al.* Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat Protoc* **11**, 1455–1476, doi:10.1038/nprot.2016.086 (2016).
14. Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445, doi:10.1016/j.cell.2010.03.030 (2010).
15. Aguilera, A. & Garcia-Muse, T. R loops: from transcription byproducts to threats to genome stability. *Mol Cell* **46**, 115–124, doi:10.1016/j.molcel.2012.04.009 (2012).
16. Allison, D. F. & Wang, G. G. R-loops: formation, function, and relevance to cell stress. *Cell Stress* **3**, 38–46, doi:10.15698/cst2019.02.175 (2019).
17. Skourti-Stathaki, K. & Proudfoot, N. J. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev* **28**, 1384–1396, doi:10.1101/gad.242990.114 (2014).
18. Sanz, L. A. & Chedin, F. High-resolution, strand-specific R-loop mapping via S9.6-based DNA-RNA immunoprecipitation and high-throughput sequencing. *Nat Protoc* **14**, 1734–1755, doi:10.1038/s41596-019-0159-1 (2019).
19. Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I. & Chedin, F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell* **45**, 814–825,

- doi:10.1016/j.molcel.2012.01.017 (2012).
20. Zatreanu, D. *et al.* Elongation Factor TFIIIS Prevents Transcription Stress and R-Loop Accumulation to Maintain Genome Stability. *Mol Cell* **76**, 57–69 e59, doi:10.1016/j.molcel.2019.07.037 (2019).
 21. Akman, G. *et al.* Pathological ribonuclease H1 causes R-loop depletion and aberrant DNA segregation in mitochondria. *Proc Natl Acad Sci U S A* **113**, E4276-4285, doi:10.1073/pnas.1600537113 (2016).
 22. Parajuli, S. *et al.* Human ribonuclease H1 resolves R-loops and thereby enables progression of the DNA replication fork. *J Biol Chem* **292**, 15216–15224, doi:10.1074/jbc.M117.787473 (2017).
 23. Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science* **362**, doi:10.1126/science.aav1898 (2018).
 24. Dravis, C. *et al.* Epigenetic and Transcriptomic Profiling of Mammary Gland Development and Tumor Models Disclose Regulators of Cell State Plasticity. *Cancer Cell* **34**, 466–482 e466, doi:10.1016/j.ccell.2018.08.001 (2018).
 25. Lee, S. O. *et al.* Inactivation of the orphan nuclear receptor TR3/Nur77 inhibits pancreatic cancer cell and tumor growth. *Cancer Res* **70**, 6824–6836, doi:10.1158/0008-5472.CAN-10-1992 (2010).
 26. Lee, S. O. *et al.* The nuclear receptor TR3 regulates mTORC1 signaling in lung cancer cells expressing wild-type p53. *Oncogene* **31**, 3265–3276, doi:10.1038/onc.2011.504 (2012).
 27. Wu, H. *et al.* Regulation of Nur77 expression by beta-catenin and its mitogenic effect in colon cancer cells. *FASEB J* **25**, 192–205, doi:10.1096/fj.10-166462 (2011).
 28. Chen, J. *et al.* NR4A transcription factors limit CAR T cell function in solid tumours. *Nature* **567**, 530–534, doi:10.1038/s41586-019-0985-x (2019).
 29. Liu, X. *et al.* Genome-wide analysis identifies NR4A1 as a key mediator of T cell dysfunction. *Nature* **567**, 525–529, doi:10.1038/s41586-019-0979-8 (2019).
 30. Tan, C. *et al.* NR4A nuclear receptors restrain B cell responses to antigen when second signals are absent or limiting. *Nat Immunol*, doi:10.1038/s41590-020-0765-7 (2020).
 31. Eberhardy, S. R. & Farnham, P. J. Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the cad promoter. *J Biol Chem* **277**, 40156–40162, doi:10.1074/jbc.M207441200 (2002).
 32. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* **13**, 720–731, doi:10.1038/nrg3293 (2012).
 33. Skourti-Stathaki, K., Kamieniarz-Gdula, K. & Proudfoot, N. J. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* **516**, 436–439, doi:10.1038/nature13787 (2014).
 34. Nojima, T. *et al.* Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161**, 526–540, doi:10.1016/j.cell.2015.03.027 (2015).

Figures

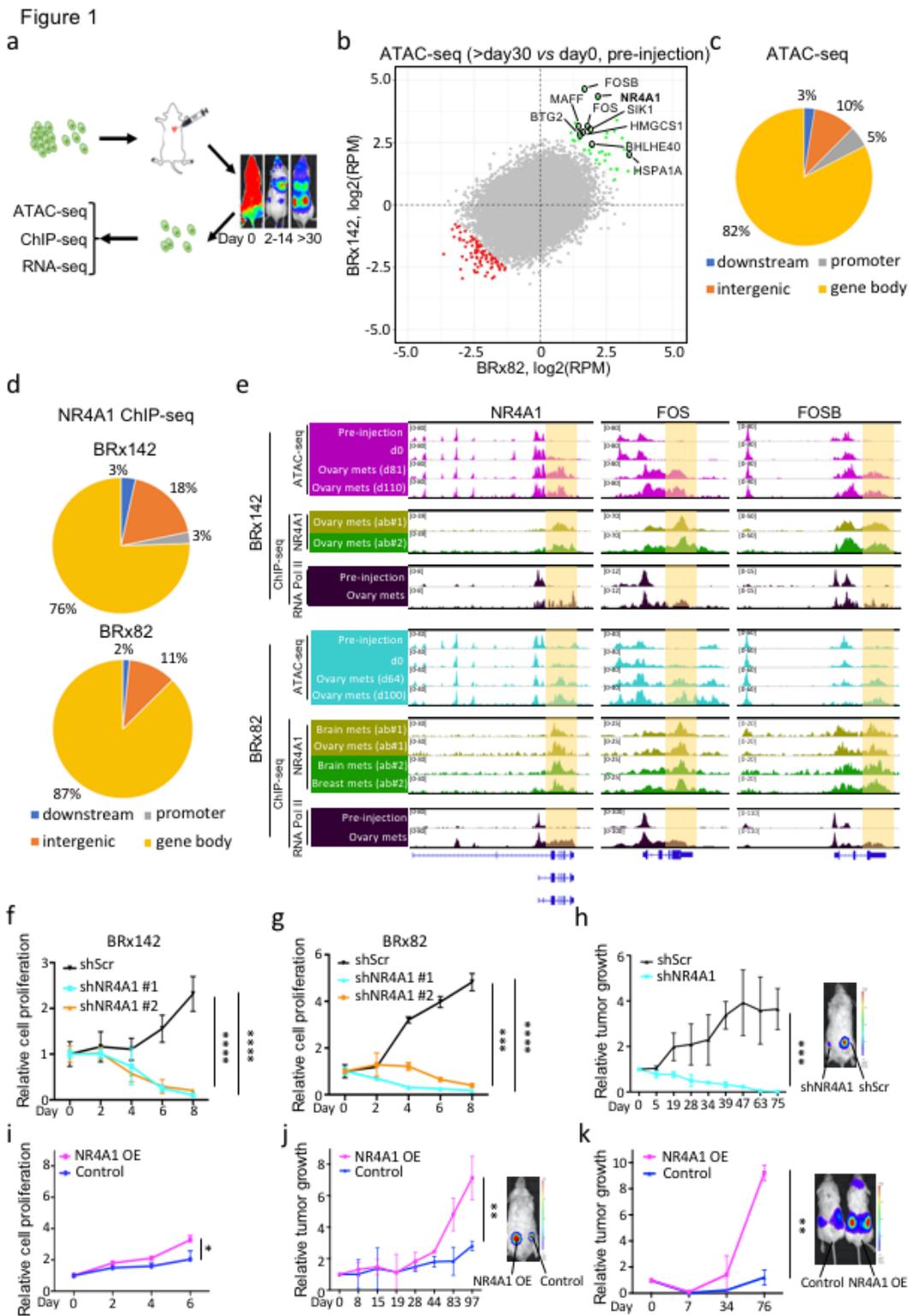


Figure 1

Temporal dynamics of chromatin accessibility in tumor cells during tumor initiation in vivo. (a) Schematic of general workflow, from intracardiac injection of cultured GFP-luciferase-tagged breast circulating tumor cells (CTCs) to serial in vivo imaging of metastases, recovery of tagged single cells from mouse tissues and genomic characterization. (b) Scatter plot showing regions gaining (green) or losing (red) chromatin accessibility (ATAC-seq) across both BRx142 and BRx82 CTC-derived metastases

(>day30 vs day0 and preinjection). Immediate Early Genes (IEGs) gaining chromatin accessibility are identified. (c) Pie chart showing the distribution of genomic regions gaining chromatin accessibility during metastasis, with marked predominance of genebody regions. (d) Pie charts showing the distribution of genomic regions of NR4A1 ChIP-seq peaks in metastatic tumors from BRx142 and BRx82 CTCs, with marked predominance of genebody regions. (e) IGV tracks showing ATAC-seq (purple or cyan), NR4A1 ChIP-seq (yellow and green) and total RNA Pol II ChIP-seq (black) signal at three IEGs, NR4A1, FOS and FOSB (upper panel: BRx142; bottom panel: BRx82). Accumulation of reads at IEG gene bodies is shown and the 3'-end is highlighted. For NR4A1 ChIP-seq, yellow and green tracks represent ChIP-seq signal generated using two different antibodies (see Methods). (f-g) In vitro cell proliferation of BRx142 (f) and BRx82 (g) CTCs with NR4A1 shRNA knockdown, compared with parental control. Statistical significance was assessed by two-tailed Student's t-Test. *** $P < 0.001$; **** $P < 0.0001$, $n = 4$. (h) Average bioluminescence signal of orthotopic tumors derived from BRx142 cells with NR4A1 knockdown (cyan) or scrambled control (black) over time. Statistical significance was assessed by two-tailed Student's t-Test. *** $P < 0.001$, $n = 6$. Representative NSG mouse with paired orthotopic tumors (shRNR4A1 right mammary gland, left image; shScr, left gland, right image) using IVIS imaging at day 75. (i) In vitro cell proliferation of BRx142 CTCs with ectopic NR4A1 overexpression (OE), compared with parental control. Statistical significance was assessed by two-tailed Student's t-Test. * $P < 0.05$, $n = 4$. (j) In vivo proliferation of BRx142 CTCs following orthotopic mammary fat pad inoculation, measured by average bioluminescence signal over time, for cells with NR4A1 overexpression (OE) versus control. Statistical significance was assessed by two-tailed Student's t-Test. ** $P < 0.01$, $n = 6$. A representative mouse bioluminescence image (day 83) is shown (NR4A1 OE cells right gland, left image; parental cells left gland, right image). (k) In vivo proliferation of BRx142 CTCs (NR4A1 OE versus control) following intracardiac inoculation, serially measured by quantitation of bioluminescence signal over time. Statistical significance was assessed by two-tailed Student's t-Test. ** $P < 0.01$, $n = 4$. Representative mice (NR4A1 OE versus control) are shown at day 76 post intracardiac injection.

Figure 2

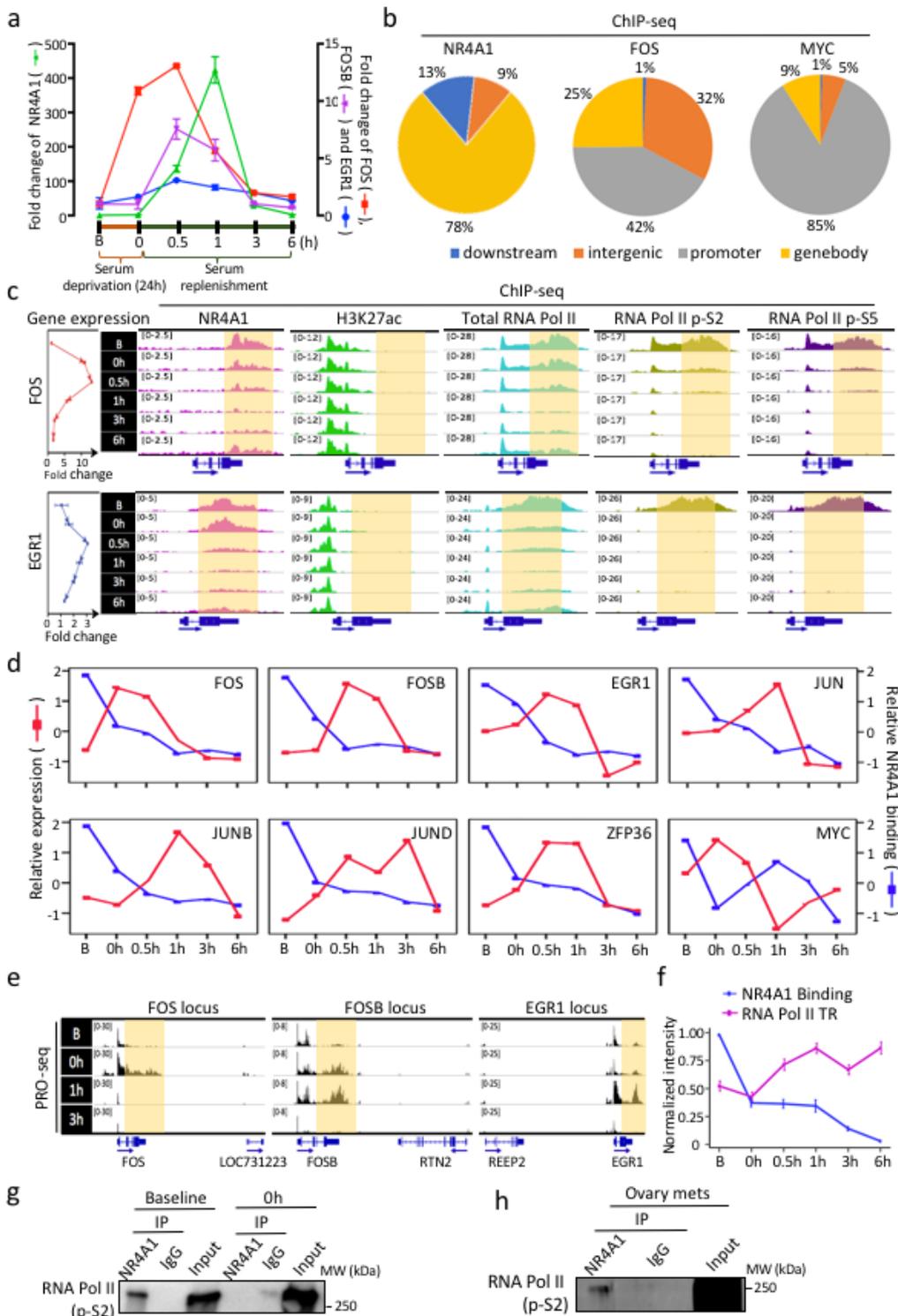


Figure 2

NR4A1 restrains RNA Polymerase II transcriptional elongation at IEG genebodies. (a) IEG induction (mRNA quantitation) in MCF10A cells grown under baseline culture conditions (“B”) and subjected to serum deprivation (0h), followed by replenishment (0.5, 1, 3 and 6h). Error bars represent SD. (b) Pie chart showing marked predominance of genebody binding by NR4A1 across the genome compared with other IEGs: FOS and MYC, which primarily bind to the promoter or intergenic regions, all under basal serum

conditions (c) IGV tracks showing ChIP binding dynamics, at serial time points after serum withdrawal and replenishment, of NR4A1, H3K27ac, total RNA Pol II, RNA Pol II phospho S2 and RNA Pol II phospho S5, at two IEG loci, FOS and EGR1. Schematic gene structure and transcriptional direction is shown below the IGV tracks. The baseline (“B”) accumulation of NR4A1 ChIP reads toward the 3’-end of FOS and EGR1 genebodies (highlighted in shaded yellow) is resolved upon serum withdrawal (0h) and replenishment (0.5, 1, 3 and 6h), overlapping in pattern with the RNA Pol II ChIP dynamics, whereas H3K27ac at gene promoters is unchanged. Serial expression of FOS and EGR1 mRNA (left panel) is inversely correlated with genebody localization by both NR4A1 and RNA Pol II. (d) Inverse correlation of NR4A1 binding at IEG genebodies (blue) versus gene expression (red) for multiple IEGs (FOS, FOSB, EGR1, JUN, JUNB, JUND, ZFP36 and MYC) at sequential timepoints following serum starvation and replenishment (e) IGV tracks showing nuclear run-on sequencing (PRO-seq) signal at IEG genebodies under baseline (“B”), serum withdrawal (0h) and serum replenishment (1, 3h). IEGs show new transcription upon starvation and refeeding, whereas their neighboring genes do not show any PRO-seq signal, demonstrating specificity. (f) RNA Pol II Traveling Ratio (TR, purple), compared with NR4A1 binding (blue) at IEG genebodies, following serum withdrawal and replenishment, demonstrating inverse correlation between NR4A1 binding and RNA Pol II activity along IEG gene bodies. Curves represent average for all NR4A1-bound genes (n=39). See Methods for TR calculation. (g-h) Co-immunoprecipitation followed by western blot showing protein association between NR4A1 and RNA Pol II (CTD phospho S2) in MCF10A cells under baseline culture conditions, but not following serum withdrawal (0h) (g), and in CTC-derived ovarian metastasis (h).

Figure 3

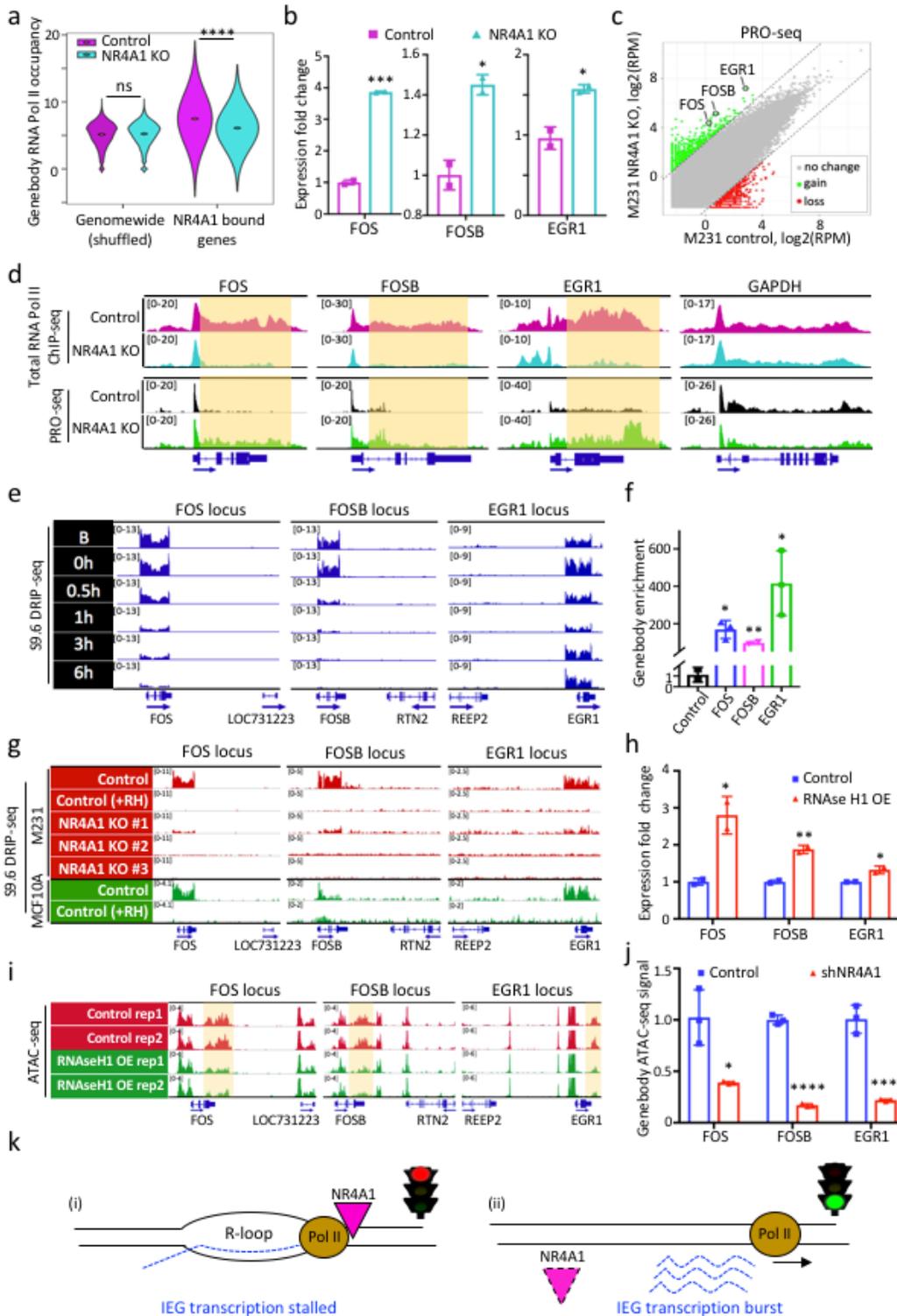


Figure 3

NR4A1 depletion enables IEG transcription through resolution of R-loops. (a) Violin plots showing reduced total RNA Pol II read densities at NR4A1-bound genebodies (n=39) in NR4A1-deleted (KO) M231 cells, compared to parental controls, without genome-wide change in RNA Pol II localization (****P<0.0001; ns, not significant, by Wilcox test). (b) Increased expression of IEGs in NR4A1-null M231 cells, versus parental controls, under baseline culture conditions. Error bar represents SD. *P<0.05; ***P<0.001, by two-

tailed Student's t-Test, $n=2$. (c) Scatter plot showing gain of PRO-seq signal indicating active nuclear run-on and RNA polymerase processing at IEGs (FOS, FOSB, EGR1) in NR4A1-null M231 cells, compared with parental controls under baseline condition. (d) IGV tracks showing increased PRO-seq signal (nuclear run-on) coincident with reduced RNA Pol II ChIP-seq signal at genebodies of IEGs (FOS, FOSB, EGR1) in NR4A1-null M231 cells, versus parental controls, under baseline culture. Genebody and 3'-UTR sequences are highlighted. GAPDH control shows minimal changes of total RNA Pol II and PRO-seq signals upon NR4A1 knockout (KO). (e) IGV tracks showing S9.6 DRIP-seq signal indicating presence of R-loops across three IEG loci in M231 cells under basal serum condition and conditions of serum withdrawal and replenishment. The neighboring gene for each IEG, shows no R-loops, demonstrating specificity of S9.6 DRIP-seq. (f) S9.6 DRIP-qPCR assay showing enrichment of R-loop signal at IEG genebody regions (FOS, FOSB, EGR1) in BRx82 CTC-derived metastases, compared with a control genebody locus (GAPDH). Error bar denotes SD. * $P<0.05$; ** $P<0.01$ by two-tailed Student's t-Test, $n=3$. (g) IGV tracks showing S9.6 DRIP-seq signal across three IEG loci in both M231 cells and MCF10A cells under baseline serum condition. The R-loop signal is virtually erased after in vitro RNaseH1 treatment (+RH), confirming DNA-RNA hybrids. S9.6 DRIP-seq signal is also absent in NR4A1-null M231 cells (KO #1-3), consistent with the role of NR4A1 in initiating R-loop formation. The neighboring gene for each IEG, shows no R-loop signal. (h) Upregulation of IEG (FOS, FOSB and EGR1) expression in M231 cells with in vivo ectopic expression of RNase H1, grown under baseline culture conditions, compared with uninfected parental controls. * $P<0.05$; ** $P<0.01$ by two-tailed Student's t-Test, $n=2$. (i) IGV tracks showing ATAC-seq signal at IEG genebody regions in MCF10A cells under basal serum condition (two biological replicates). IEG genebody and 3'-UTR regions are highlighted in shaded yellow showing the reduction of ATAC-seq signal after ectopically expressing RNase H1 in vivo (RNaseH1 OE). (j) Bar graph showing the reduction of genebody ATAC-seq signal after shRNA-mediated NR4A1 knockdown in MCF10A cells grown under basal serum conditions. Error bar denotes SD. * $P<0.05$; *** $P<0.001$; **** $P<0.0001$ by two-tailed Student's t-Test, $n=3$. (k) Schematic model showing that (i) NR4A1 normally restrains RNA Pol II transcriptional elongation at IEG genebody, halting transcription and generating R-loops, and (ii) stress-induced displacement of NR4A1 or deletion of NR4A1 enables RNA Pol II elongation, triggering a burst of IEG expression.

Figure 4

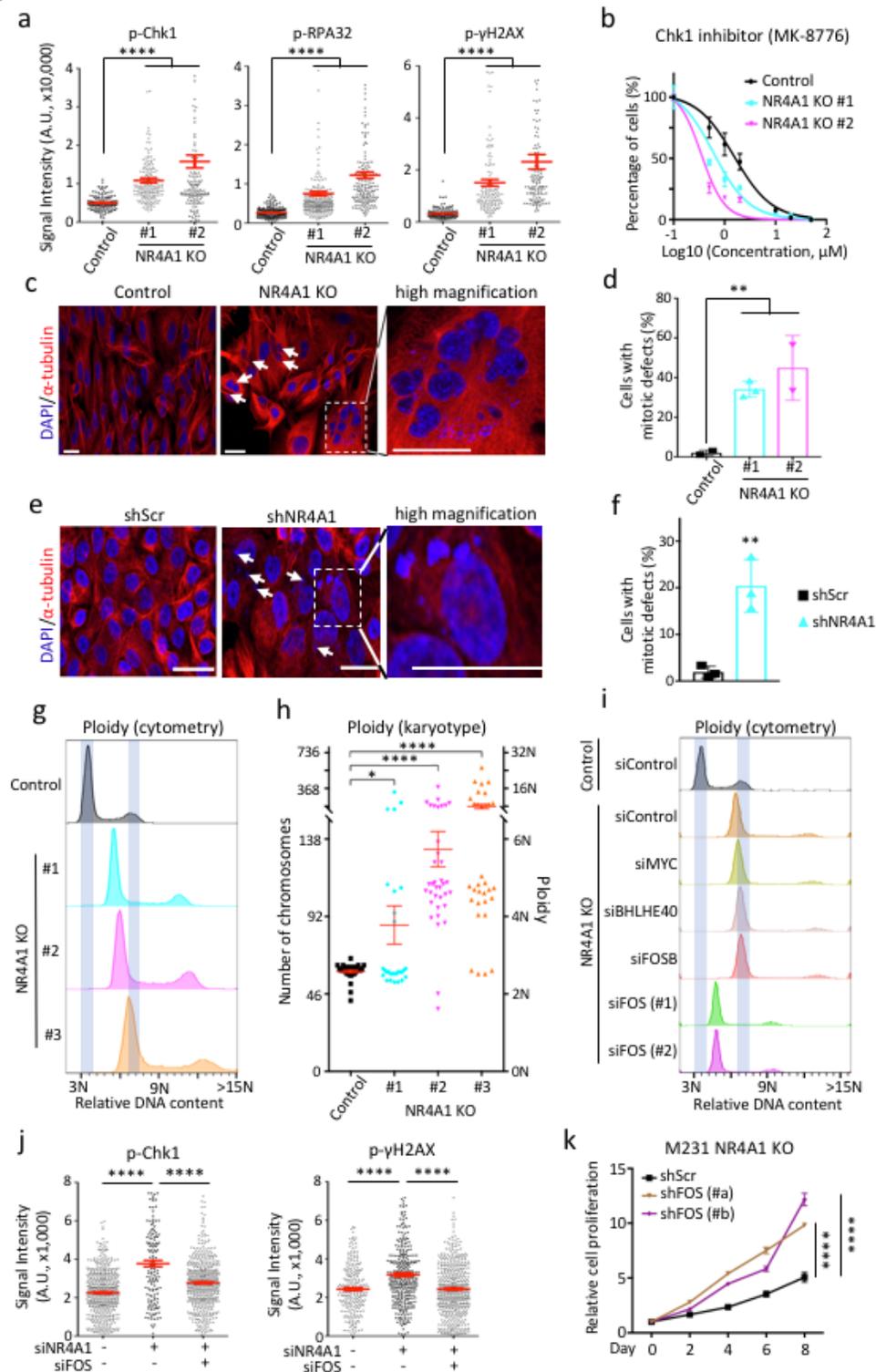


Figure 4

NR4A1 deletion triggers FOS-mediated replication stress and genomic instability. (a) Quantification of individual cells with positive phospho-Chk1, phospho-RPA32 and phospho-γH2AX foci in two NR4A1-deleted (KO) M231 clones, versus control parental cells. Error bar in red denotes SEM. *** $P < 0.001$ by two-tailed Student's t-Test, $n > 100$. Image quantification was carried out by ImageJ software. (b) Increased sensitivity of two independent NR4A1-null M231 clones to the Chk1 inhibitor MK-8776, compared with

parental control. Error bar indicates SD, n=4. Y-axis represents relative cellular viability; x-axis shows log₁₀ drug concentrations. (c) Representative images of dual α -tubulin and nuclear (DAPI) immunofluorescence staining, showing increase of multinucleated cells and multinuclei in NR4A1-null M231 cells, which are absent in parental controls. A higher magnification of the inset is shown on the right. Scalebar, 50 μ M. Arrows indicate some of the cells with mitotic defects. (d) Bar graph showing the percentage of cells with gross mitotic defects in two M231 NR4A1-null clones (#1-2), compared with parental control. Error bar represents SD. **P<0.01 by two-tailed Student's t-Test, n=3. (e) Dual staining for α -tubulin and DAPI showing multiple micronuclei in MCF10A cells following shRNA-mediated suppression of NR4A1. A representative image is shown at low and high magnification. No mitotic defects are evident in a representative image of MCF10A cells infected with control shScr. Bar, 50 μ M. (f) Bar graph showing percentage of MCF10A cells with mitotic defects after shRNA-mediated NR4A1 knockdown. Error bar indicates SD. *P<0.05 by two-tailed Student's t-Test, n=3. (g) DNA ploidy analysis by flow cytometry of three independent NR4A1-null M231 clones (#1-3), compared with parental controls, showing increased aneuploidy. X-axis represents relative DNA content in cells after FxCycle Violet DNA staining. (h) Single-cell karyotyping showing large-scale gain of chromosomes in three NR4A1-null M231 clones, compared with parental controls. Error bar denotes SEM. *P<0.05; ****P<0.0001 by two-tailed Student's t-Test, n>25. (i) DNA ploidy analysis by flow cytometry showing resolution of aneuploidy in NR4A1-null M231 cells, following suppression of FOS expression using siRNA (technical replicates: #1-2) versus scrambled control. No such effect is observed following knockdown of other IEGs bound by NR4A1, including MYC, BHLHE40, or FOSB. X-axis represents relative DNA content in cells after FxCycle Violet DNA staining. (j) Quantification of cells with positive phospho-Chk1 and phospho- γ H2AX foci in M231 cells at baseline, showing increase following siRNA knockdown of NR4A1, and rescue following simultaneous siRNA-mediated knockdown of NR4A1 and FOS. Error bar in red denotes SEM. ****P<0.0001 by two-tailed Student's t-Test, n>100. Image quantification was carried out by ImageJ software. (k) Rescue of in vitro cell proliferation defect in NR4A1-null M231 cells, following knockdown of FOS using two different shRNA constructs (#a-b). ****P<0.0001 by two-tailed Student's t-Test, n=4.

Figure 5

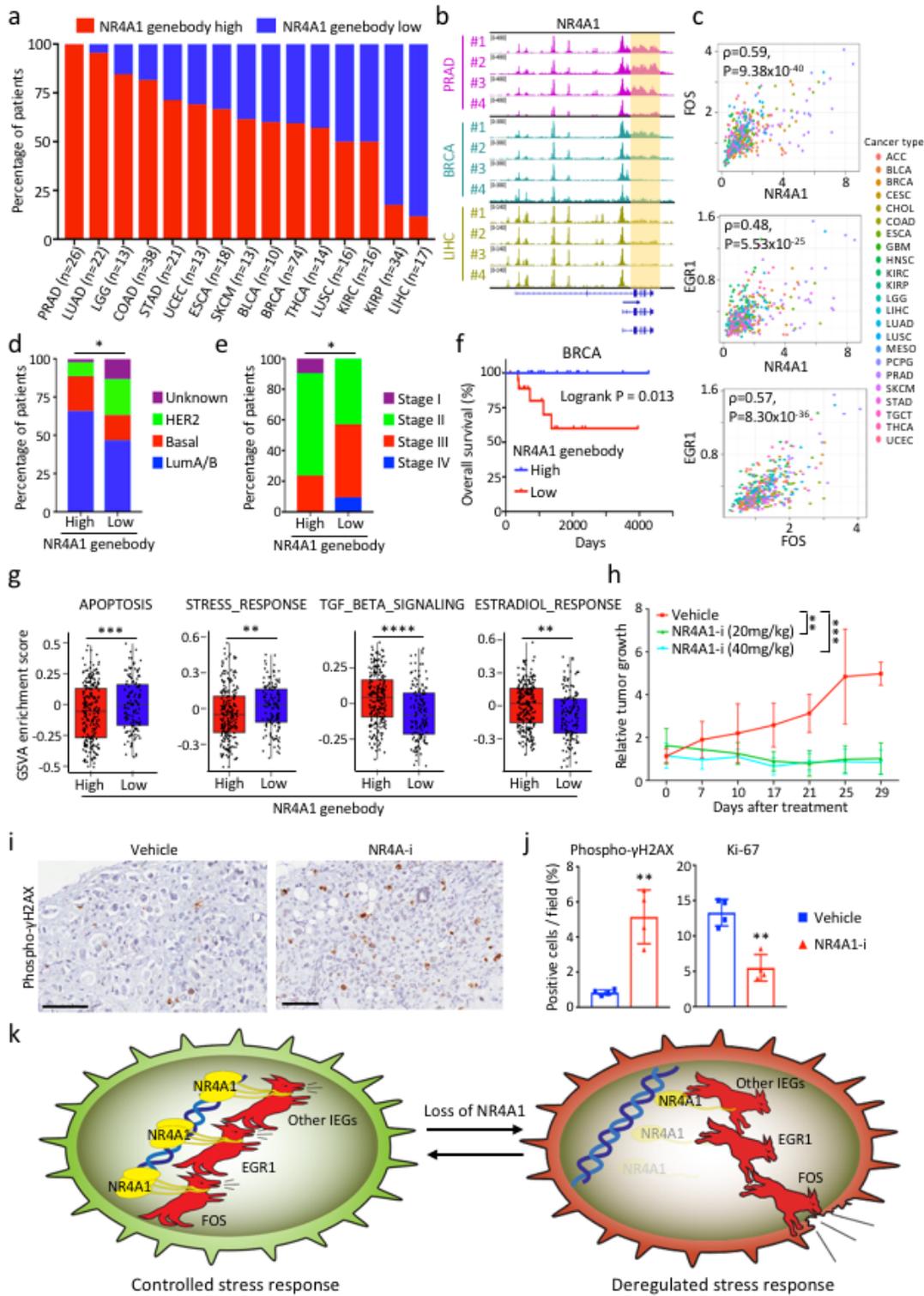


Figure 5

Prevalence of NR4A1 and IEG genebody chromatin accessibility in primary human cancers. (a) Percentage of patients with primary tumor showing high (red bars) or low (blue bars) ATAC-seq signal at the NR4A1 genebody across TCGA cancer types. Only cancer types with more than 10 ATAC-seq samples are shown. Patients are classified into high and low groups according to the NR4A1 genebody/promoter ATAC-seq signal ratio (see methods). Primary ATAC-seq data from ref23. (b) Representative IGV tracks

showing normalized ATAC-seq signal at NR4A1 genebody (highlighted in shaded yellow) in prostate (PRAD), breast (BRCA) and liver (LIHC) cancers, derived from ref23. (c) Scatter plots showing pairwise correlation of genebody ATAC-seq signal among three IEGs (NR4A1, FOS and EGR1) across all TCGA cancer types. Spearman's correlation coefficient (ρ) and p-value are shown. (d) Bar graph showing the percentage of breast cancer patients with different molecular subtypes (PAM50) showing high or low NR4A1 genebody ATAC-seq signal. $**P < 0.01$ by Fisher exact test. (e) Bar graph showing lower pathologic tumor stages (AJCC) in breast cancers with high NR4A1 genebody ATAC-seq signal. $*P < 0.05$ by Fisher exact test. (f) Kaplan-Meier plot showing improved overall survival in breast cancer patients with high NR4A1 genebody ATAC-seq signal. $P = 0.0013$, calculated by logrank test. (g) Box plots showing GSVA analysis of different pathways in TCGA primary tumors with high ($n = 237$) or low ($n = 140$) NR4A1 genebody ATAC-seq signal, indicating reduced apoptosis and stress signaling, and increased TGF- β and Estrogen Receptor signaling in tumors with high NR4A1 genebody ATAC-seq signal. $**P < 0.01$; $***P < 0.001$; $****P < 0.0001$ assessed by two-sided Welch's t-test. (h) Suppression of in vivo orthotopic tumorigenesis by patient-derived breast BRx142 CTCs in NSG mice, following treatment with NR4A1 inhibitor (NR4A1-i; 20 mg/kg/day or 40 mg/kg/day; see methods). Average bioluminescence signal is shown over time. $**P < 0.01$; $***P < 0.001$ assessed by two-sided Student's t-test, $n = 6$. (i) Immunohistochemistry staining (IHC) images showing elevated phospho- γ H2AX staining (brown) in tumors after treatment of NR4A1-i (Day 29). Scalebar, 100 μ M. (j) Quantification of cells positive for DNA damage marker phospho- γ H2AX or proliferation marker Ki-67 per field in tumor histologic sections after in vivo treatment of mice with either vehicle or NR4A1-i. Error bar represents SD. $**P < 0.01$ by two-tailed Student's t-Test, $n = 4$. (k) The model of "Let slip the IEGs," paraphrasing "Cry 'Havoc!' and let slip the dogs of war" (Shakespeare, Julius Caesar), illustrating how IEG expression is normally kept leashed by NR4A1, but the stress-mediated release of NR4A1 from IEG genebodies removes the brake on their transcriptional elongation, resulting in a massive burst of IEG expression that can lead to cell death. Cancer cells with preserved IEG genebody open chromosome domains may depend on this NR4A1 activity to survive oncogenic replication stress.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [GuoNaturesupplementaryinformation2X0200925.docx](#)
- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable3.xlsx](#)