Transcription-dependent regulation of replication dynamics modulates genome stability (85/90)

Marion Blin[1,2,3,7]*, *Benoît Le Tallec*[1,2,3,8]*+, *Viola Naehse*[1,2,3,9], *Mélanie Schmidt*[4], *Gael A. Millot*[5], *Marie-Noëlle Prioleau*[6] *and Michelle Debatisse*[1,2,3,10+]

1 Institut Curie, Centre de Recherche, Paris, France.

2 Université Pierre & Marie Curie, Sorbonne Universités, Paris, France.

3 CNRS UMR 3244, Paris, France.

4 Institut Gustave Roussy, CNRS UMR 8200, Villejuif, France.

5 Institut Pasteur, Bioinformatics and Biostatistics Hub, C3BI, USR 3756 IP CNRS, Paris, France.

6 Institut Jacques Monod, CNRS UMR 7592, Université Paris Diderot, Paris, France.

7 Present address: INSERM, Aix Marseille Université, CNRS, Institut Paoli-Calmettes, CRCM, Marseille, France.

8 Present address: Ecole Normale Supérieure, Institut de Biologie de l'ENS (IBENS), INSERM U1024, CNRS UMR 8197, Paris, France.

9 Present address: Department of Molecular Cell Biology, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo, Norway.

10 Present adress : Institut Gustave Roussy, CNRS UMR 8200, Villejuif, France.

*: co-first authors

+: co-corresponding authors

Replication stress is a primary threat to genome stability and has been implicated in tumorigenesis[1,2]. Common fragile sites (CFSs) are loci hypersensitive to replication stress[3] and are hotspots for chromosomal rearrangements in cancers[4]. CFSs replicate late in S-phase[3], are cell-type dependent[4-6] and nest within very large genes[4,7-9]. The mechanisms responsible for CFS instability are still discussed, notably the relative impact of transcription-replication conflicts[7,8,10] *versus* their low density in replication initiation events[5,6]. Here we address the relationships between transcription, replication, gene size and instability by manipulating the transcription of three endogenous large genes, two in chicken and one in human cells. Remarkably, moderate transcription destabilises large genes whereas high transcription levels alleviate their instability. Replication dynamics analyses showed that transcription quantitatively shapes the replication program of large genes, setting both their initiation profile and their replication timing as well as regulating internal fork velocity. Noticeably, high transcription levels advance the replication time of large genes from late to mid S-phase, which most likely gives cells more time to complete replication before mitotic entry. Transcription can therefore contribute to maintaining the integrity of some difficult-to-replicate loci, challenging the dominant view that it is exclusively a threat to genome stability.

It is largely agreed that CFSs tend to remain incompletely replicated until mitosis upon replication stress. Incompletely replicated regions are processed by specific endonucleases promoting mitotic DNA synthesis and sister chromatid separation, eventually at the cost of chromosomal rearrangements[11-15]. Two main mechanisms have been suggested to explain this delayed replication completion. One postulates that secondary DNA structures[10] or transcription-dependent replication barriers, notably R-loops[7,8,10], lead to fork stalling and collapse. The other proposes that replication of the core of the CFSs by long-travelling forks due to their paucity in initiation events is specifically delayed upon fork slowing[5,6]. Here we

directly addressed the impact of transcription on the replication program and fragility of very large genes. For this purpose, we manipulated the transcription of such genes in chicken DT40 cells, where targeted DNA modification by homologous recombination is very efficient, and in human colon carcinoma HCT116 cells using the CRISPR-Cas9 technique.

We first focused on *DMD*, the largest annotated avian gene (Fig. 1). *DMD* extends over 996 kb and is neither transcribed (Fig. 1a) nor fragile (Fig. 1b) in DT40 cells. Determination of the replication timing showed that *DMD* replicates late (S4 fraction) in S-phase (Fig. 1e). We activated its transcription by inserting a tetracyclin-inducible promoter (Tet-promoter) on both alleles (Supplementary Fig. 1). A low level of *DMD* transcription from the Tet-promoter was detected in $DMD^{Tet/Tet}$ cells grown without tetracycline (Fig. 1a). Importantly, fluorescent *in situ* hybridization (FISH) analysis of breaks induced in *DMD* by the DNA polymerase inhibitor aphidicolin revealed that the gene became fragile in these cells (Fig. 1b). Addition of tetracycline increased both *DMD* mRNA levels (Fig. 1a) and *DMD* instability (Fig. 1b). Our results therefore strongly support a major role of transcription in fragility, further illustrating that transcription constitutes an intrinsic source of genomic instability[16-18].

We then examined the replication program of *DMD* in wild-type (WT) cells and in $DMD^{Tet/Tet}$ cells grown in the presence of tetracycline. We used DNA-combing to visualize neo-synthesized DNA along single molecules spanning *DMD*, which permitted us to measure the speed of individual forks and to map initiation and termination events along the locus[5] (Supplementary Figs 2, 3a and Supplementary Table 1). Initiation and termination events appeared randomly distributed along *DMD* in WT cells (Fig. 1c). In striking contrast, initiation events accumulated in the Tet-promoter region when *DMD* was transcribed (Fig. 1d). This initiation zone was followed by a large region depleted of initiation events extending over the 5' part of the gene. Initiation events also clustered in a second, ≈300 kb-large zone located in the 3' half of active *DMD*. As expected, modifications of the initiation

profile resulted in a redistribution of termination events. Active *DMD* remained late-replicating (Fig. 1f), confirming that transcription *per se* is not sufficient to impose early replication[19,20]. Thus, our results show that transcription modulates origin distribution, in line with recent analyses[21-23], and account for the frequent localization of origins near transcription start sites[24,25]. In addition, activation of *DMD* transcription specifically increased fork speed along the *DMD* locus (Fig. 1g). One hypothesis is that transcription could regulate chromatin permissiveness to fork progression. Consistently, chromatin factors like the histone chaperone FACT (Facilitates Chromatin Transcription), which participates in the disassembly of the nucleosomes upstream of the RNA polymerase, influence the rate of fork progression on *in vitro* reconstituted chromatin templates[26].

We next modulated the transcription of the 616 kb-long *CCSER1* gene in DT40 cells (Fig. 2). *CCSER1* is transcribed (Fig. 2a) and is one of the most fragile regions of the DT40 genome upon aphidicolin treatment (Fig. 2b and [4]). We enhanced *CCSER1* transcription by placing both alleles of the gene under the control of the strong chicken β-actin promoter (Supplementary Fig. 4). Unexpectedly, the ensuing ≈20 fold increase in *CCSER1* transcription (Fig. 2a and Supplementary Fig. 4f) was accompanied by a dramatic reduction of aphidicolin-induced *CCSER1* instability (Fig. 2b). To understand why, we compared *CCSER1* replication in WT and CCSER1$^{\beta a/\beta a}$ cells (Fig. 2c-g, Supplementary Fig. 3b and Supplementary Table 1). A strong initiation zone overlapped *CCSER1* promoter region in WT cells, with some additional initiation events apparently randomly distributed along the gene (Fig. 2c). Importantly, *CCSER1* overexpression further favoured initiation in the promoter region at the expense of the rest of the gene (Fig. 2d). In addition, termination events clustered in the third quarter of *CCSER1* in CCSER1$^{\beta a/\beta a}$ cells (Fig. 2d), suggesting that CCSER1$^{\beta a}$ allele was mainly replicated by forks proceeding inward from the initiation zone located upstream of the β–actin promoter and from a second initiation region located 3' of *CCSER1*. Together with

DMD, our results therefore demonstrate that transcription quantitatively dictates the replication initiation program of active large genes, as well as the ensuing termination profile. In agreement with the results obtained for *DMD*, enhancing *CCSER1* transcription also significantly increased fork velocity inside the locus (Fig. 2g).

Remarkably, *CCSER1* overexpression advanced the replication time of the 5' part of the gene, which shifted from mid (S2/S3 fractions) to early S-phase (S1/S2 fractions) (Fig. 2f). It also markedly advanced the replication time of the 3' end of *CCSER1* and its flanking region (Fig. 2f). These data are consistent with those obtained by combing (Fig. 2d), reciprocally validating each other and further supporting that transcription, at least at a sufficient level, stimulates initiation both upstream and downstream of large genes. One intriguing possibility is that the coordinated timing shift observed in 5' and 3' of *CCSER1* upon overexpression might be mediated by long-range chromatin interactions. Advanced initiation at both ends of *CCSER1* collectively led to a replication timing shift of the whole *CCSER1* locus in CCSER1$^{\beta a/\beta a}$ cells, mainly from late (S4 fraction) to mid-late (S3/S4 fractions) S-phase (Fig. 2e, f). These results suggest that the stability of CCSER1$^{\beta a}$ alleles stems from the advancement of replication timing induced by the strong increase in transcription, which gives cells more time to complete replication before entering mitosis. Together with the data obtained for *DMD*, they also suggest that only high transcription levels may be capable of advancing the replication timing of large genes.

To test this hypothesis, we used the β-actin promoter in place of the Tet-promoter to increase *DMD* transcription (Supplementary Fig. 5a-f). We failed to recover homozygous cells but heterozygous DMD$^{+/\beta a}$ cells were alive and exhibited no growth defect (Supplementary Fig. 5g). We therefore compared these cells to heterozygous DMD$^{+/Tet}$ cells treated with tetracycline (Fig. 3). *DMD* transcription was ≈10 times higher in DMD$^{+/\beta a}$ than in DMD$^{+/Tet}$ cells (Fig. 3a and Supplementary Fig. 5h) yet, in line with *CCSER1* results, *DMD* was two-

fold less fragile upon aphidicolin exposure (Fig. 3b). As anticipated, we detected a strong shift to earlier replication timing at the β-actin promoter insertion site, from S4 to S2/S3 fractions (Fig. 3c, d).

Finally, we sought to extend our conclusions to FRA3B, the archetypal human CFS nested in the 1.5-Mb-long *FHIT* tumour suppressor gene[27] (Fig. 4). We used the CRISPR-Cas9 system to up-regulate *FHIT* transcription in HCT116 epithelial cells, where FRA3B is the most active CFS[4] (Supplementary Fig. 6). *FHIT* promoter region was substituted on both alleles with a cassette containing the strong human EF1α promoter (Supplementary Fig. 6), which resulted in a ≈20 fold increase in *FHIT* mRNA levels (Fig. 4a). This increase elicited a massive reduction of FRA3B instability upon aphidicolin treatment (Fig. 4b) accompanied by a shift to earlier replication of the *FHIT* locus, from late (S3/S4 fractions) to mid S-phase (S2/S3 fractions) (Fig. 4c, d). Thus, data from human cells are perfectly in agreement with those from DT40 cells, further substantiating that transcription directly and quantitatively influences the replication time of large genes and that advancing replication timing protects active large genes from fragility. Interestingly, the 5' and 3' ends of *FHIT* replicated earlier than its central part in WT cells (Fig. 4c). This profile is accentuated in FHIT$^{EF1\alpha/EF1\alpha}$ cells (Fig. 4d), reminiscent of the replication timing profile of CCSER1$^{\beta a}$ alleles (Fig. 2f). These results confirm that high levels of transcription favour advanced initiation of replication both 5' and 3' of active large genes, contributing to an earlier replication of the entire locus. Noticeably, it has been reported that the insertion in the genome of DT40 cells of the strong β-actin promoter upstream of the 423 bp-long blasticidin resistance gene has little impact on the replication timing of the targeted chromosomal region[19], suggesting that transcription-dependent shift in timing may be limited to genes over a certain size.

We show here that the degree of instability of large genes upon replication stress relies on their transcription level in an atypical way, with low transcription being more deleterious than

high transcription levels. Strikingly, while cause-and-effect relationships between transcriptional regulation and replication timing have fuelled an intense debate[28], we demonstrate that transcription regulates replication timing in a quantitative rather than qualitative manner. Noticeably, high levels of transcription advance the replication time from late to mid S-phase of all the three large genes that we studied and protect them from replication stress-induced breaks, which establishes a causal role for late replication timing in CFS instability. The high transcriptional-level dependent shift in replication timing most likely gives cells more time to complete replication prior to the onset of mitosis. Importantly, it was observed in multiple cell types that active large genes are generally transcribed at low levels[8,9] and, accordingly, tend to replicate late during S-phase[8]. Our results are therefore in line with recent genome-wide analyses concluding that CFSs specifically correspond to large, transcribed, late-replicating genes and also explain why not all transcribed large genes are fragile[4,29].

The role of transcription in origin distribution observed for both *DMD* and *CCSER1* prompted us to propose that initiation-poor regions and associated long–travelling forks observed in CFSs[3,5,6] actually originate from and reflect the transcription of the cognate large genes. Whether CFS instability results from transcription-driven paucity in initiation events or transcription-dependent formation of barriers impeding fork progression or both phenomena remains to be determined.

In conclusion, while transcription is considered today as a primary threat to genome stability, our study reveals complex relationships between transcription, replication and chromosome fragility. The multifaceted impact of transcription on replication calls for a careful assessment of the mutational consequences of transcriptional changes, especially in cancer cells where up-regulation of transcription is suspected to be a major contributor of oncogene-induced replication stress[30].

References

1 Gaillard, H., Garcia-Muse, T. & Aguilera, A. Replication stress and cancer. *Nature reviews. Cancer* **15**, 276-289, doi:10.1038/nrc3916 (2015).

2 Macheret, M. & Halazonetis, T. D. DNA replication stress as a hallmark of cancer. *Annual review of pathology* **10**, 425-448, doi:10.1146/annurev-pathol-012414-040424 (2015).

3 Debatisse, M., Le Tallec, B., Letessier, A., Dutrillaux, B. & Brison, O. Common fragile sites: mechanisms of instability revisited. *Trends in genetics : TIG* **28**, 22-32, doi:10.1016/j.tig.2011.10.003 (2012).

4 Le Tallec, B. *et al.* Common fragile site profiling in epithelial and erythroid cells reveals that most recurrent cancer deletions lie in fragile sites hosting large genes. *Cell reports* **4**, 420-428, doi:10.1016/j.celrep.2013.07.003 (2013).

5 Letessier, A. *et al.* Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature* **470**, 120-123, doi:10.1038/nature09745 (2011).

6 Le Tallec, B. *et al.* Molecular profiling of common fragile sites in human fibroblasts. *Nature structural & molecular biology* **18**, 1421-1423, doi:10.1038/nsmb.2155 (2011).

7 Helmrich, A., Ballarino, M. & Tora, L. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Molecular cell* **44**, 966-977, doi:10.1016/j.molcel.2011.10.013 (2011).

8 Wilson, T. E. *et al.* Large transcription units unify copy number variants and common fragile sites arising under replication stress. *Genome research* **25**, 189-200, doi:10.1101/gr.177121.114 (2015).

9 Wei, P. C. *et al.* Long Neural Genes Harbor Recurrent DNA Break Clusters in Neural Stem/Progenitor Cells. *Cell* **164**, 644-655, doi:10.1016/j.cell.2015.12.039 (2016).

10 Madireddy, A. *et al.* FANCD2 Facilitates Replication through Common Fragile Sites. *Molecular cell* **64**, 388-404, doi:10.1016/j.molcel.2016.09.017 (2016).

11 Naim, V., Wilhelm, T., Debatisse, M. & Rosselli, F. ERCC1 and MUS81-EME1 promote sister chromatid separation by processing late replication intermediates at common fragile sites during mitosis. *Nature cell biology* **15**, 1008-1015, doi:10.1038/ncb2793 (2013).

12 Ying, S. *et al.* MUS81 promotes common fragile site expression. *Nature cell biology* **15**, 1001-1007, doi:10.1038/ncb2773 (2013).

13 Minocherhomji, S. *et al.* Replication stress activates DNA repair synthesis in mitosis. *Nature* **528**, 286-290, doi:10.1038/nature16139 (2015).

14 Bhowmick, R., Minocherhomji, S. & Hickson, I. D. RAD52 Facilitates Mitotic DNA Synthesis Following Replication Stress. *Molecular cell* **64**, 1117-1126, doi:10.1016/j.molcel.2016.10.037 (2016).

15 Sotiriou, S. K. *et al.* Mammalian RAD52 Functions in Break-Induced Replication Repair of Collapsed DNA Replication Forks. *Molecular cell* **64**, 1127-1134, doi:10.1016/j.molcel.2016.10.038 (2016).

16 Hamperl, S. & Cimprich, K. A. Conflict Resolution in the Genome: How Transcription and Replication Make It Work. *Cell* **167**, 1455-1467, doi:10.1016/j.cell.2016.09.053 (2016).

17 Garcia-Muse, T. & Aguilera, A. Transcription-replication conflicts: how they occur and how they are resolved. *Nature reviews. Molecular cell biology* **17**, 553-563, doi:10.1038/nrm.2016.88 (2016).

18 Hamperl, S., Bocek, M. J., Saldivar, J. C., Swigut, T. & Cimprich, K. A. Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. *Cell* **170**, 774-786 e719, doi:10.1016/j.cell.2017.07.043 (2017).

19 Hassan-Zadeh, V. *et al.* USF binding sequences from the HS4 insulator element impose early replication timing on a vertebrate replicator. *PLoS biology* **10**, e1001277, doi:10.1371/journal.pbio.1001277 (2012).

20 Rivera-Mulia, J. C. *et al.* Dynamic changes in replication timing and gene expression during lineage specification of human pluripotent stem cells. *Genome research* **25**, 1091-1103, doi:10.1101/gr.187989.114 (2015).

21 Petryk, N. *et al.* Replication landscape of the human genome. *Nature communications* **7**, 10208, doi:10.1038/ncomms10208 (2016).

22 Gros, J. *et al.* Post-licensing Specification of Eukaryotic Replication Origins by Facilitated Mcm2-7 Sliding along DNA. *Molecular cell* **60**, 797-807, doi:10.1016/j.molcel.2015.10.022 (2015).

23 Powell, S. K. *et al.* Dynamic loading and redistribution of the Mcm2-7 helicase complex through the cell cycle. *Embo J* **34**, 531-543, doi:10.15252/embj.201488307 (2015).

24 Prioleau, M. N. & MacAlpine, D. M. DNA replication origins-where do we begin? *Genes & development* **30**, 1683-1697, doi:10.1101/gad.285114.116 (2016).

25 Aladjem, M. I. & Redon, C. E. Order from clutter: selective interactions at mammalian replication origins. *Nature reviews. Genetics* **18**, 101-116, doi:10.1038/nrg.2016.141 (2017).

26 Kurat, C. F., Yeeles, J. T., Patel, H., Early, A. & Diffley, J. F. Chromatin Controls DNA Replication Origin Selection, Lagging-Strand Synthesis, and Replication Fork Rates. *Molecular cell* **65**, 117-130, doi:10.1016/j.molcel.2016.11.016 (2017).

27 Waters, C. E., Saldivar, J. C., Hosseini, S. A. & Huebner, K. The FHIT gene product: tumor suppressor and genome "caretaker". *Cellular and molecular life sciences : CMLS* **71**, 4577-4587, doi:10.1007/s00018-014-1722-0 (2014).

28 Rivera-Mulia, J. C. & Gilbert, D. M. Replication timing and transcriptional control: beyond cause and effect-part III. *Current opinion in cell biology* **40**, 168-178, doi:10.1016/j.ceb.2016.03.022 (2016).

29 Miron, K., Golan-Lev, T., Dvir, R., Ben-David, E. & Kerem, B. Oncogenes create a unique landscape of fragile sites. *Nature communications* **6**, 7094, doi:10.1038/ncomms8094 (2015).

30 Kotsantis, P. *et al.* Increased global transcription activity as a mechanism of replication stress in cancer. *Nature communications* **7**, 13087, doi:10.1038/ncomms13087 (2016).

Acknowledgments

We thank S. Lambert, O. Hyrien, F. De Carli, M. Hennion, L. Lacroix and V. Besic for critical reading of the manuscript. The authors would like to acknowledge the Cell and Tissue Imaging Platform - PICT-IBiSA (member of France–Bioimaging) of the Genetics and Developmental Biology Department (UMR3215/U934) of Institut Curie for help with light microscopy, the Flow Cytometry Platform Imagoseine of Institut Jacques Monod, Université Paris Diderot, and the Imaging and Cytometry Platform (PFIC) of Institut Gustave Roussy for assistance with cell sorting. M. D. team is supported by the Agence Nationale de la Recherche (ANR-13-BSV6-0008-01/FRA-Dom), the Association pour la Recherche sur le Cancer (Subvention Libre Sl220130607073) and the Institut National du Cancer (INCa subvention 2013-103). M. N. P. team is supported by the Association pour la Recherche sur le Cancer (Labellisation PGA120150202272) and the Agence Nationale de la Recherche (ANR-15-CE12-0004-01). M. B. was supported by fellowships from the Ministère de l'Enseignement Supérieur et de la Recherche and the Ligue contre le cancer.

Methods

Cell culture. DT40 cells were grown in RPMI 1640 GlutaMAX medium (Gibco) with 10% fetal bovine serum (Biowest), 2% chicken serum (Gibco) and 0.1 mM β-mercaptoethanol (Millipore). A tetracycline-free fetal bovine serum (Biowest) was used when appropriate. HCT116 human colon carcinoma cells were cultured in McCOY'S 5A medium (Gibco) with 10% fetal calf serum and 20 mM L-glutamine (Gibco) except for replication timing experiments for which they were cultured in DMEM (4.5 g/L D-Glucose, L-glutamine) (Gibco) with 10% fetal calf serum and 1 mM sodium pyruvate. All media were supplemented with 100 $U.mL^{-1}$ penicillin and 100 $\mu g.mL^{-1}$ streptomycin (Gibco). Cells were grown at 37°C, 20% O_2, 5% CO_2. DT40 and HCT116 cells and derivative clones were routinely confirmed to be negative for mycoplasma contamination.

Tet-ON inducible system. A Tet-ON system based on the T-REx System (Invitrogen) was used to induce the transcription of *DMD* upon addition of the tetracycline antibiotic. Briefly, the tetracycline-inducible promoter (Tet-promoter) used consists of a CMV promoter into which 2 copies of the Tet-operator 2 (TetO) sequence have been inserted in tandem. In the absence of tetracycline, transcription is repressed by the high affinity binding of the Tet-repressor (TetR) to the TetO. Addition of tetracycline results in the release of TetO by the TetR and derepression of the promoter. The Tet-promoter was amplified from a pcDNA4/TO vector (Invitrogen). The TetR was expressed from a pcDNA6/TR vector (Invitrogen).

Constructs. Targeting constructs for homologous recombination in DT40 cells were created from LoxP vectors[1] using standard molecular biology and cloning protocols. ≈2 kb-long 5' and 3' homology regions were amplified from DT40 genomic DNA with primers listed in

Supplementary Table 2. The chicken β-actin promoter was synthesized by GeneScript. All constructs were checked for mutations by sequencing. Guide RNAs targeting *FHIT* promoter (gRNA1: GCCAAATGCCATGTGGGTGC; gRNA2: TCAATTTAGATTTCGGCTTC) were designed using the gRNA design tool from DNA2.0. gRNA/Cas9 plasmid containing sequences for the two gRNAs and wtCas9 was synthesized by DNA2.0. ≈1 kb-long 5' and 3' homology regions flanking the CRISPR/Cas9 cutting sites were amplified from RP11-137N22 BAC with primers listed in Supplementary Table 2 and cloned into the HR710PA-1 plasmid (System Biosciences) on each side of the EF1α promoter-hygromycin selection cassette. Detailed cloning procedures are available upon request.

Transfection and genotyping. 10^7 exponentially growing DT40 cells were electroporated with 35 μg of the linearized construct using a Biorad electroporator set at 25 μF and 550 V for targeted integrations, 960 μF and 250 V for the random insertion of the pcDNA6/TR vector. Recombinant clones were selected with either 21.75 $\mu g.mL^{-1}$ blasticidin, 0.5 $\mu g.mL^{-1}$ puromycin or 450 $\mu g.mL^{-1}$ zeocin, identified by PCR using LA Taq DNA Polymerase (Takara) and confirmed by Southern blot. At least two positive clones were randomly selected and amplified for further studies. Primers used for genotyping are listed in Supplementary Table 2. Probes used for Southern blot were amplified from the targeting constructs using primers listed in Supplementary Table 2. The 2-log DNA ladder (NEB) was used as molecular-weight size marker for agarose gel electrophoresis except in Supplementary Fig. 1f where the GeneRuler 1 kb Plus DNA ladder (ThermoFisher Scientific) was used.

For transient transfection, $3x10^6$ exponentially growing DT40 cells were transfected with 5 μg of a reporter plasmid expressing the luciferase gene under the control of the Tet-promoter using the Amaxa Nucleofector system (T solution, program B-023).

HCT116 cells were co-transfected with 1 µg gRNA/Cas9 plasmid and 1µg HR710PA-1-*FHIT* plasmid using FuGENE HD Transfection Reagent (Promega) according to the manufacturer's instructions. Recombinant cells were selected with 200 $\mu g.mL^{-1}$ hygromycin B (Invitrogen) 48 h post-transfection. 2 µM ganciclovir was added to the culture medium 7 days post-transfection to counter-select cells containing randomly integrated HR710PA-1-*FHIT* plasmid. After a further 12 days, single clones were selected and amplified. Positive clones were identified by PCR and confirmed by sequencing.

Luciferase reporter assay. DT40 clonal cell lines expressing the Tet-repressor were tested for their ability to induce transcription of the luciferase reporter gene upon addition of tetracycline, which was quantified by the light emitted during an enzymatic reaction of bioluminescence using the Dual-Glo Luciferase Assay System (Promega) according to the manufacturer's instructions. Luminescence was measured using a luminometer 24 h after addition of 1 $\mu g.mL^{-1}$ tetracycline (SIGMA) to the culture medium.

Tetracycline-induced transcriptional activation of *DMD*. Transcriptional activation of *DMD* was achieved with a 24 h treatment of $DMD^{Tet/Tet}$ and $DMD^{+/Tet}$ cells with 1 $\mu g.mL^{-1}$ tetracycline. A low transcriptional background from the Tet-promoter is detected even without tetracycline (Figs 1a and 3a), as already observed[2].

Floxed cassette excision. The DT40 cell line used in this study contains a stably integrated MerCreMer plasmid[1]. The protein is inactive in the absence of 4-hydroxytamoxifen (4-OHT) due to the hormone binding domains of the Mutated estrogen receptor (Mer) fused either side of the Cre recombinase. For excision of floxed cassettes, cells were cultured for 24 h with 0.5 µM 4-OHT then distributed into 96-well flat-bottom microtiter plates at a concentration of 1,

348 3 and 30 viable cells per well. Isolated colonies were tested for cassette excision by loss of
349 antibiotic resistance and PCR.
350
351 **mRNA quantification.** DT40 cells were collected during their exponential growth phase.
352 Total RNA was extracted using miRNeasy Mini Kit (Qiagen) according to the manufacturer's
353 instructions and treated with DNase (Roche). 1 μg of total RNA was reverse-transcribed using
354 the iScript cDNA Synthesis Kit (Biorad). Real-time quantitative PCR reactions were set up
355 using GoTaq qPCR Master Mix (Promega) and run on a LightCycler 480 II (Roche). Each
356 reaction was performed in triplicate. DNA contamination was quantified in reverse
357 transcriptase free reactions. Primer sequences are listed in Supplementary Table 3. Each
358 primer pair spanned one intron to avoid unwanted amplification of genomic DNA. The
359 efficiency of each primer pair was tested by performing qPCR with the same protocol on
360 increasing dilutions of cDNA (from 1/10 to 1/10,000 dilutions) and calculated using the
361 coefficient of amplification. The efficiency of each pair being similar, it is possible to directly
362 compare the amount of each amplification product. mRNA levels were quantified relative to
363 β-actin (*ACTB*) mRNA. For $DMD^{Tet/Tet}$, $DMD^{+/Tet}$ and $DMD^{+/\beta a}$ cells, mRNA levels were
364 quantified before the excision of the floxed selection cassettes. For each construct, mRNA
365 quantification experiments were performed using at least two different clonal cell lines.
366 Total RNA from HCT116 cells was isolated using RNeasy Mini Kit (Qiagen) according to the
367 manufacturer's instructions. cDNA was synthesized with Maxima First Strand cDNA
368 Synthesis Kit for RT-qPCR (ThermoFisher Scientific). Real-time quantitative PCR was
369 performed using Luminaris Color HiGreen qPCR Master Mix, low ROX (ThermoFisher
370 Scientific) on a 7500 Real-Time PCR System (Applied Biosystems). Primer sequences are
371 listed in Supplementary Table 3. *FHIT* mRNA levels were quantified relative to RNA
372 polymerase II subunit F (*POLR2F*), ribosomal protein L11 (*RPL11*) and cyclophylin B

373 (*PPIB*) mRNA. mRNA quantification experiments were performed using one FHIT$^{EF1\alpha/EF1\alpha}$
374 clonal cell line.
375
376 **Quantification of 5-Ethynyl uridine (EU) incorporation into nascent RNAs**. 10^7
377 exponentially growing DT40 cells were treated for 1 h with 0.5 mM EU. EU-labelled RNA
378 was captured using the Click-iT Nascent RNA Capture kit (Invitrogen) according to the
379 manufacturer's instructions. Briefly, 1 µg of total RNA was biotinylated thanks to the click
380 reaction between EU and azide-modified biotin, then purified using streptavidin magnetic
381 beads. Nascent RNAs were reverse-transcribed on the beads using Superscript kit
382 (Invitrogen). Real-time quantitative PCR reactions were performed as described above for
383 DT40 samples using intra-intronic primers listed in Supplementary Table 3. All primers had
384 similar efficiencies. RNA levels were calculated relative to β-actin mRNA quantities.
385
386 **Fluorescent *in situ* hybridization (FISH) on metaphase chromosomes.** DT40 and HCT116
387 cells were grown for 16 h with 0.6 and 0.15 µM aphidicolin, respectively. Metaphase spreads
388 were prepared according to standard cytogenetic procedures after a 3 h treatment with 0.1
389 $\mu g.mL^{-1}$ colcemid for DT40 cells and a 2 h treatment with 100 nM nocodazol for HCT116
390 cells. For untreated DT40 cells, metaphase spreads were prepared after a 1.5 h treatment with
391 0.1 $\mu g.mL^{-1}$ colcemid.
392 For DT40 FISH analyses, probes correspond to adjacent 5 kb-long PCR products spread over
393 ≈50 kb delimiting *DMD*, *CCSER1* or *PARK2* genes (see probe coordinates in Supplementary
394 Table 4); probes were amplified from DT40 genomic DNA with primers listed in
395 Supplementary Table 4. Probes were labelled either with biotin using the BioPrime DNA
396 labelling system (Invitrogen) or with digoxigenin using the DIG DNA labelling mix (Roche)
397 and subsequently purified on Illustra ProbeQuant G-50 Micro Columns (GE Healthcare).

FISH was performed essentially as described[3] without the proteinase K step and using 75 ng of each probe. Chicken Hybloc DNA (Amplitech) was used as a repetitive sequence competitor DNA. Immunodetection was performed by successive incubations in the following reagents: (i) for biotinylated probes, (1) Alexa Fluor 555-conjugated streptavidin (Invitrogen), (2) biotin-conjugated rabbit anti-streptavidin (Rockland Immunochemicals), (3) Alexa Fluor 555-conjugated streptavidin (Invitrogen) (ii) for digoxigenin-labelled probes, (1) FITC-conjugated mouse anti-digoxin (Jackson ImmunoResearch), (2) Alexa Fluor 488-conjugated goat anti-mouse (Invitrogen). Chromosomes were counterstained with 49,6-diamidino-2-phenylindole (DAPI) (Vectashield mounting medium for fluorescence with DAPI; Vector Laboratories) and metaphases were observed by fluorescence microscopy. Aphidicolin-induced breaks at the 679.5 kb-long *PARK2* gene were used as a control to demonstrate that modulation of *DMD* or *CCSER1* transcription specifically impacts the fragility of those genes and that the limited perturbations of cell physiology observed in CCSER1$^{\beta a/\beta a}$ cells (Supplementary Fig. 4d, e) do not perturb break induction by aphidicolin. For each construct, FISH experiments were performed using at least two different clonal cell lines.

FISH analysis of FRA3B fragility in HCT116 cells was performed using 50 ng of biotinylated RP11-641C1, RP11-32J15 and RP11-147N17 BACs selected from the human genome project RP11 library. Immunodetection was performed by successive incubations in Alexa Fluor 488-conjugated streptavidin (Invitrogen) and biotin-conjugated rabbit anti-streptavidin (Rockland Immunochemicals). 15 µg Cot-1 DNA (Invitrogen) were used as a repetitive sequence competitor DNA. FISH experiments were performed using one FHIT$^{EF1\alpha/EF1\alpha}$ clonal cell line.

Bivariate Fluorescence-Activated Cell Sorting (FACS) analysis. Cell cycle analyses were performed as follows. Exponentially growing DT40 cells were pulse-labelled for 15 min with 30 µM bromodeoxyuridine (BrdU) and fixed in ethanol. After partial denaturation of DNA

following HCl/pepsin treatment, immunodetection was performed by incubations with rat anti-BrdU (Bio-Rad, formerly AbD Serotec) then Alexa Fluor 488-conjugated chicken anti-rat (Invitrogen). DNA was counterstained with propidium iodide in the presence of 25 $\mu g.mL^{-1}$ RNase. Samples were analyzed using a BD Biosciences LSRII flow cytometer with BD FACSDiva software. Data were processed using FlowJo v8.7.3. Gating strategy is illustrated in Supplementary Fig. 7.

Cell growth and doubling time. The cumulative growth curves of the DT40 cell lines used in this study were determined after normalization for dilution at each subculture. Doubling times were estimated based on the growth curves. Indeed, for exponentially growing cells, the final number of cells (N) is given by the formula N = initial number of cells x $2^{\text{number of doublings}}$, where the number of doublings corresponds to the duration of culture divided by the doubling time (T); it follows that T = [duration of culture*log(2)]/[log(final number of cells)-log(initial number of cells)].

Replication timing analyses. Timing analyses in DT40 cells were made as previously described[4]. Briefly, exponentially growing DT40 cells were pulse-labelled for 1 h with 3 mM BrdU, fixed in ethanol and stored at -20°C overnight. Fixed cells were re-suspended in 1X PBS with 50 $\mu g.mL^{-1}$ propidium iodide and 1 $mg.mL^{-1}$ RNase, and incubated for 30 min at room temperature. Cells were sorted by flow cytometry based on their nuclear content using a BD Biosciences INFLUX cell sorter with the BD FACS™ Sortware software (v 1.0.0.650). S-phase was divided into four fractions from early to late S-phase named S1 to S4. 50,000 cells were sorted in each fraction. DNA was extracted by phenol/chloroform, sonicated to obtain fragments between 500-1000 bp in size, and BrdU-labelled DNA was immunoprecipitated using mouse anti-BrdU antibody (BD Biosciences). Real-time

quantitative PCR was performed using the Roche LightCycler 2.0 detection system with the Absolute QPCR-SYBR Green mix (ThermoFisher Scientific). For each reaction, amplification of the purified BrdU-labelled DNA was performed in duplicate. As mitochondrial DNA replicates throughout the cell cycle and should be equally represented in every fraction, the amount of immunoprecipitated DNA in each S-phase fraction was normalized by the abundance of mitochondrial DNA measured using a specific primer pair (MIT). Quality control experiments were performed to confirm enrichment of known early-, mid- and late-replicated loci in the expected fractions (Supplementary Table 5). In heterozygous cells, primer pairs overlapping the site of insertion and next to the site of insertion of the Tet- or β-actin promoters were used to detect the timing of the wild type allele ("without" primers) and both alleles ("both" primers), respectively. A primer pair specific to the transgene ("with" primers) was used to analyze the timing of the modified allele. For each construct, replication timing analyses were performed on two different clonal cell lines.

For replication timing analyses of *FHIT*, exponentially growing HCT116 cells were pulse-labelled for 1 h with 50 μM BrdU, fixed in ethanol and incubated overnight at -20°C in the presence of 15 $\mu g.mL^{-1}$ Hoescht 33342 (ThermoFisher Scientific). Fixed cells were re-suspended in 1X PBS and cells were sorted by flow cytometry based on their nuclear content using a BD Biosciences INFLUX cell sorter. S-phase was divided into four fractions, from early to late S-phase (S1 to S4). To check the quality of sorted HCT116 fractions, the post-sort cells, already stained with Hoescht 33342, were directly re-analyzed by flow cytometry. DNA was isolated using Maxwell RSC Blood DNA kit (Promega) according to the manufacturer's instructions, and sonicated. For each fraction, immunoprecipitation of BrdU-labelled DNA was performed on 5 μg of DNA using mouse anti-BrdU antibody (BD Biosciences). Immunoprecipitated DNA was further extracted by phenol/chloroform, and real-time quantitative PCR was performed using QuantiNova SYBR Green PCR kit (Qiagen)

473 on a 7300 Real-Time PCR System (Applied Biosystems). Replication timing analyses were
474 performed using one FHIT$^{EF1\alpha/EF1\alpha}$ clonal cell line.
475 All primer pairs used for replication timing analyses are listed in Supplementary Table 6.
476
477 **DNA-combing.** Neo-synthesized DNA was labelled as described[5] with the following
478 changes: exponentially growing DT40 cells were pulse-labelled for 20 min with 20 μM
479 iododeoxyuridine (IdU) followed by a 20 min pulse with 100 μM chlorodeoxyuridine (CldU),
480 then by a 5 min chase with 1 mM thymidine. Genomic DNA was extracted and combing was
481 performed on silanized coverslips prepared by plasma cleaning and liquid-phase silanization
482 as described[6,7] using a Genomic Vision apparatus.
483
484 **FISH on combed DNA and immunofluorescence detection of neo-synthesized DNA.**
485 Morse-codes were designed as described[8] to specifically identify DNA molecules spanning
486 the *DMD* or *CCSER1* loci. Morse-code probes correspond to a collection of ≈5 kb-long PCR
487 products spread all over the locus of interest, separated by precise distances and divided into
488 voluntarily different and asymmetrical patterns to correctly orient the DNA fibre. Morse-
489 codes for *DMD* and *CCSER1* detection are made of 32 probes and 27 probes, respectively.
490 PCR products were prepared and labelled with biotin as described above for FISH on
491 metaphases. Primer pairs used are listed in Supplementary Table 7. Hybridization of the
492 probes was carried out as described previously[5]. Immunodetection was performed by
493 successive incubations in the following reagents: (1) Alexa Fluor 488-conjugated streptavidin
494 (Invitrogen), (2) biotin-conjugated rabbit anti-streptavidin (Rockland Immunochemicals), (3)
495 Alexa Fluor 488-conjugated streptavidin (Invitrogen), mouse anti-BrdU (BD Biosciences) and
496 rat anti-BrdU (Bio-Rad, formerly AbD Serotec), (4) biotin-conjugated rabbit anti-streptavidin
497 (Rockland Immunochemicals), Alexa Fluor 350-conjugated goat anti-mouse (Invitrogen) and

Alexa Fluor 594-conjugated donkey anti-rat (Invitrogen), (5) Alexa Fluor 488-conjugated streptavidin (Invitrogen), Alexa Fluor 350-conjugated donkey anti-goat (Invitrogen) and mouse anti-single stranded DNA (Millipore), (6) Cy5.5-conjugated goat anti-mouse (Abcam) and (7) Cy5.5-conjugated donkey anti-goat (Abcam). For bulk genome analyses, immunodetection was performed as follows: (1) FITC-conjugated mouse anti-BrdU (BD Biosciences) and rat anti-BrdU (Bio-Rad, formerly AbD Serotec), (2) Alexa Fluor 488-conjugated goat anti-mouse (Invitrogen) and Alexa Fluor 555-conjugated goat anti-rat (Invitrogen), (3) mouse anti-single stranded DNA (Millipore), (4) Cy5.5-conjugated goat anti-mouse (Abcam) and (5) Cy5.5-conjugated donkey anti-goat (Abcam). Antibody incubations, washes and slide mounting were performed as reported previously[5].

Image acquisition. Images were acquired on a motorized XY stage of an Axio Imager Z2 (Carl Zeiss) or a DM6000 B (Leica) epifluorescence microscope connected to a CoolSNAP HQ2 charge-coupled device camera (Roper Scientific) and run by Metamorph software (Molecular Devices). A X100 objective was used for imaging methaphase chromosomes and a X63 objective was used for imaging combed DNA fibres. For images of the *DMD* and *CCSER1* loci, two overlays of images were set up for each microscope field. The first one combined the IdU/CldU and FISH signals to identify the fibres of interest, replicating or not. The second overlay combined the IdU, FISH and DNA signals to determine the length of the DNA fibre bearing the Morse-code. DNA counterstaining was systematically used to ensure (i) that several fibres are not overlapping (ii) that replication signals belong to the same fibre and (iii) that replication signals are intact.

Signal treatment and statistical analyses. Measurement of DNA fibre and replication tract length was performed on imaged DNA molecules using Adobe Photoshop CS5.1. DNA fibres

bearing Morse-code signals were modelled and aligned on a schematic representation of the *DMD* or *CCSER1* loci using Adobe Illustrator CS5.1. Graphical representations of *DMD* and *CCSER1* were drawn based on the locus length, the theoretical ≈2 kb/µm stretching factor of the combing apparatus indicated by the manufacturer, and the resolution of the camera at a magnification of X63 (0.1024 µm/pixel). For instance, the 996,168 bp *DMD* gene corresponded to 996.168/2/0.1024=4864.1 pixels. Morse-code probes were used to calculate the actual stretching factor of each DNA molecule by comparing the length of the fibre imaged with the microscope with the theoretical length defined according to the position of the probes. Overall, we found a mean stretching factor of 1.93 kb/µm, which is extremely close to the expected value of 2 kb/µm. Still, the stretching factor of individual fibres fluctuated from 1.7 to 2.4 kb/µm. Therefore, to prevent that variations in the stretching of the DNA molecules during the combing step introduce a bias in tract measurement, the lengths of IdU and CldU tracts were normalized by the stretching factor of the DNA fibre on which they were located. For bulk genome analyses, the lengths of IdU and CldU tracts were normalized by the mean stretching factor calculated when analyzing *DMD* or *CCSER1* loci using the exact same sample and coverslip batch. Replication fork speed was then estimated on individual forks as (i) the ratio $(l_{Idu}+l_{CldU})/(t_{IdU}+t_{CldU})$ for forks displaying an intact IdU tract flanked on one side by an intact CldU tract, (ii) the ratio l_{Idu}/t_{IdU} for forks with an intact IdU signal flanked on one side by a broken CldU signal and (iii) the ratio l_{Cldu}/t_{CldU} for forks with an intact CldU signal flanked on one side by a broken IdU signal, with l_{IdU} and t_{IdU} being the measured length (in kb) and labelling time (in min) for IdU incorporation, respectively, and l_{CldU} and t_{CldU} the corresponding parameters for CldU incorporation. For all experiments, $t_{IdU}=t_{CldU}=20$ min. Replication signal integrity was ascertained by DNA counterstaining. For fork progression analyses in *DMD* and *CCSER1*, only forks overlapping at least partially these loci were taken into account. Coordinates and length of DNA molecules and replication

548 signals were compiled using Microsoft Excel for Mac 2011. Statistical comparisons of fork
549 speed distributions were assessed with the nonparametric Mann-Whitney-Wilcoxon test (two-
550 tailed) using GraphPad Prism 6 (GraphPad Software). No assumptions or corrections were
551 made. Statistical significance was set to $p \leq 0.05$. All DNA fibres for CCSER1$^{\beta a}$ allele
552 originate from one biological sample, DNA combing results for WT *CCSER1* and *DMD* loci
553 compile DNA fibres from two distinct biological WT samples and DNA combing results for
554 DMD$^{Tet/Tet}$ cells compile DNA fibres from two clonal cell lines. Replication analysis of WT
555 and modified *CCSER1* and *DMD* loci by DNA combing was performed once.
556
557 **Coverage profile.** Graphical representations of "Total DNA" and "DNA with replication"
558 coverages of *DMD* and *CCSER1* were made with R[9] using custom scripts adapted from[10].
559
560 **Genomic coordinates and genome annotations.** Coordinates are given according to the
561 ICGSC/galGal4 chicken genome or the GRCh37/hg19 human genome assemblies. *DMD* is
562 the largest annotated gene of the chicken genome both in the latest RefSeq and Ensembl gene
563 annotations available when this manuscript was written.
564
565 **Code availabilty.** Custom R scripts are available upon request.
566
567 **Data availabilty.** The data that support the findings of this study are available upon request.
568
569 **References**
570
571 31 Arakawa, H., Lodygin, D. & Buerstedde, J. M. Mutant loxP vectors for selectable
572 marker recycle and conditional knock-outs. *BMC biotechnology* **1**, 7 (2001).

573 32 Yin, D. X., Zhu, L. & Schimke, R. T. Tetracycline-controlled gene expression system
574 achieves high-level and quantitative control of gene expression. *Analytical*
575 *biochemistry* **235**, 195-201, doi:10.1006/abio.1996.0112 (1996).
576 33 Smith, K. A., Gorman, P. A., Stark, M. B., Groves, R. P. & Stark, G. R. Distinctive
577 chromosomal structures are formed very early in the amplification of CAD genes in
578 Syrian hamster cells. *Cell* **63**, 1219-1227 (1990).
579 34 Hassan-Zadeh, V. *et al.* USF binding sequences from the HS4 insulator element
580 impose early replication timing on a vertebrate replicator. *PLoS biology* **10**, e1001277,
581 doi:10.1371/journal.pbio.1001277 (2012).
582 35 Anglana, M., Apiou, F., Bensimon, A. & Debatisse, M. Dynamics of DNA replication
583 in mammalian somatic cells: nucleotide pool modulates origin choice and interorigin
584 spacing. *Cell* **114**, 385-394 (2003).
585 36 Michalet, X. *et al.* Dynamic molecular combing: stretching the whole human genome
586 for high-resolution studies. *Science* **277**, 1518-1523 (1997).
587 37 Labit, H. *et al.* A simple and optimized method of producing silanized surfaces for
588 FISH and replication mapping on combed DNA fibers. *BioTechniques* **45**, 649-652,
589 654, 656-648 (2008).
590 38 Lebofsky, R., Heilig, R., Sonnleitner, M., Weissenbach, J. & Bensimon, A. DNA
591 replication origin interference increases the spacing between initiation events in
592 human cells. *Molecular biology of the cell* **17**, 5337-5345, doi:10.1091/mbc.E06-04-
593 0298 (2006).
594 39 The R core team. *R: A Language and Environment for Statistical Computing*. R
595 Foundation for Statistical Computing (2017).
596 40 De Carli, F., Gaggioli, V., Millot, G. A. & Hyrien, O. Single-molecule, antibody-free
597 fluorescent visualisation of replication tracts along barcoded DNA molecules. *The*

598 *International journal of developmental biology* 60, 297-304,
599 doi:10.1387/ijdb.160139oh (2016).

Figure 1

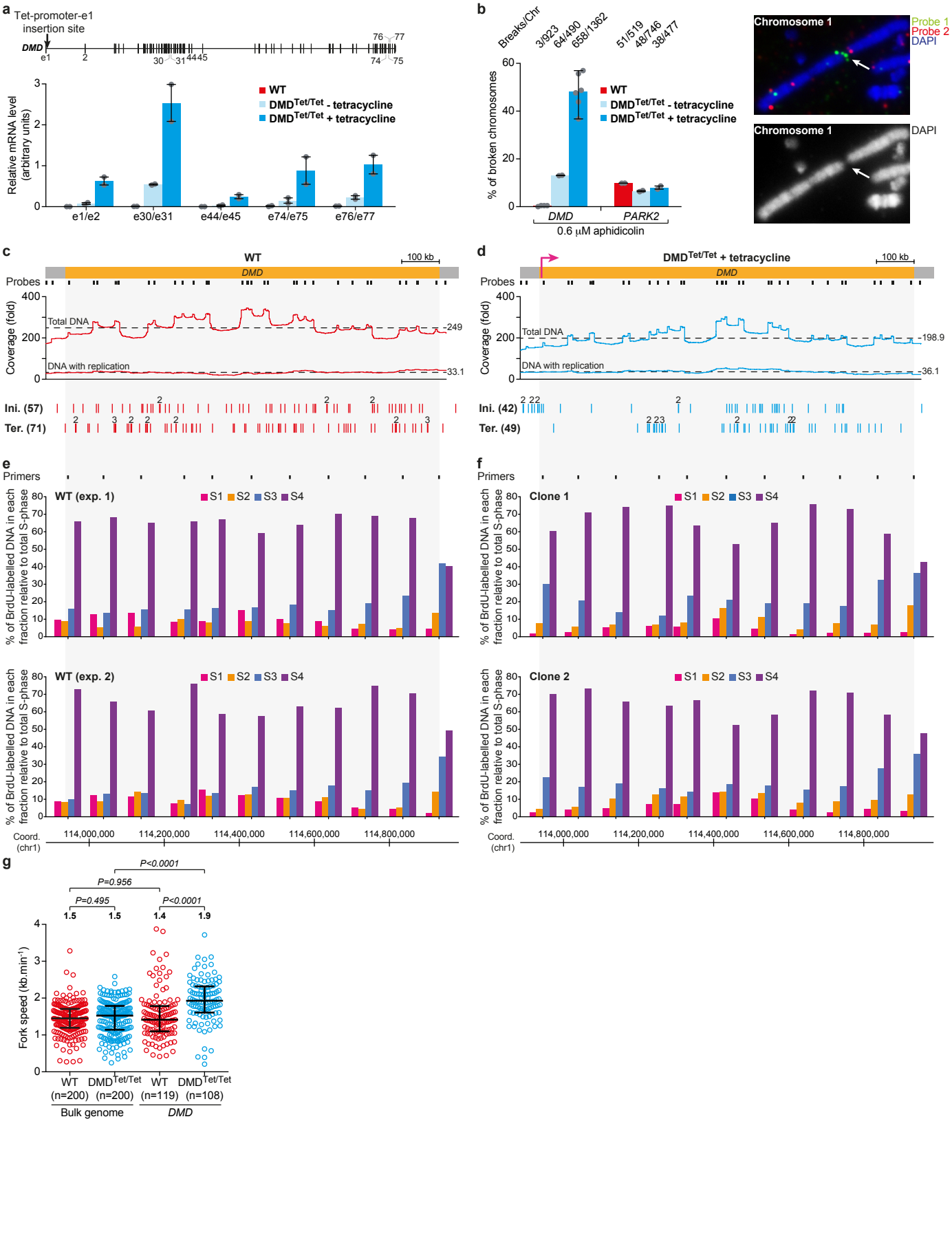
a
Tet-promoter-e1 insertion site
DMD
e1
2
30
31
44
45
74
75
76
77
Relative mRNA level (arbitrary units)
WT
DMDTet/Tet - tetracycline
DMDTet/Tet + tetracycline
e1/e2
e30/e31
e44/e45
e74/e75
e76/e77
b
Breaks/Chr
3/923
64/490
658/1362
51/519
48/746
38/477
% of broken chromosomes
DMD
PARK2
0.6 µM aphidicolin
Chromosome 1
Probe 1
Probe 2
DAPI
c
WT
100 kb
Probes
Coverage (fold)
Total DNA
249
DNA with replication
33.1
Ini. (57)
Ter. (71)
d
DMDTet/Tet + tetracycline
198.9
36.1
Ini. (42)
Ter. (49)
e
Primers
WT (exp. 1)
WT (exp. 2)
S1
S2
S3
S4
% of BrdU-labelled DNA in each fraction relative to total S-phase
Coord. (chr1)
114,000,000
114,200,000
114,400,000
114,600,000
114,800,000
f
Clone 1
Clone 2
g
P<0.0001
P=0.956
P=0.495
P<0.0001
1.5
1.5
1.4
1.9
Fork speed (kb.min-1)
WT (n=200)
DMDTet/Tet (n=200)
WT (n=119)
DMDTet/Tet (n=108)
Bulk genome
DMD

Figure 1. Impact of *DMD* transcription activation on its replication and fragility.

a, *DMD* mRNA levels relative to β-actin mRNA in WT and DMD$^{Tet/Tet}$ cells (median, extreme values and individual data points). Tested exonic junctions are indicated. Top: Map of *DMD* with its exons and the position of the Tet-promoter-*DMD* exon 1 insertion site.

b, Left panel: Aphidicolin-induced breaks at *DMD* and *PARK2* in WT and DMD$^{Tet/Tet}$ cells (median, extreme values and individual data points; aggregate numbers are presented on top). Breaks at the 679.5 kb-long *PARK2* gene were used as a control. Right panel: Example of two-color FISH with probes flanking *DMD*. Reverse-DAPI staining is also shown. The arrow indicates the position of a break. We also verified that *DMD* is not fragile in the absence of aphidicolin in WT and DMD$^{Tet/Tet}$ cells (see Supplementary Fig. 1i).

c, d, Mapping of initiation and termination events along the *DMD* locus in WT cells (**c**) and DMD$^{Tet/Tet}$ cells with tetracycline (**d**). From top to bottom: *DMD* locus (yellow box. A pink arrow represents the active Tet-promoter in DMD$^{Tet/Tet}$ cells) with the name of the cell line indicated above; Morse-code probes (black boxes) used in DNA-combing experiments to identify *DMD*; Total DNA and DNA with replication coverages of the locus (local peaks in total DNA coverage map to Morse-code probes; see Supplementary Figs 2, 3. Dotted lines represent mean coverages, indicated on the right); Map of initiation and termination events (the total number of events is indicated on the left. Figures above some lines indicate colocalized events).

e, f, Replication timing profile of *DMD* in WT cells (**e**) and DMD$^{Tet/Tet}$ cells with tetracycline (**f**). BrdU pulse-labelled cells were sorted into four S-phase fractions and neo-synthesized DNA was quantified by real-time PCR using the indicated primers (black boxes). Two experiments are shown; for DMD$^{Tet/Tet}$ cells, replication timing analyses performed on two different clonal cell lines are presented.

624 **g,** Replication fork progression in WT and $DMD^{Tet/Tet}$ cells with tetracycline in the bulk
625 genome or in *DMD*. Median with interquartile range (horizontal black lines), P-value and
626 number of forks measured (n) are indicated.

Figure 2

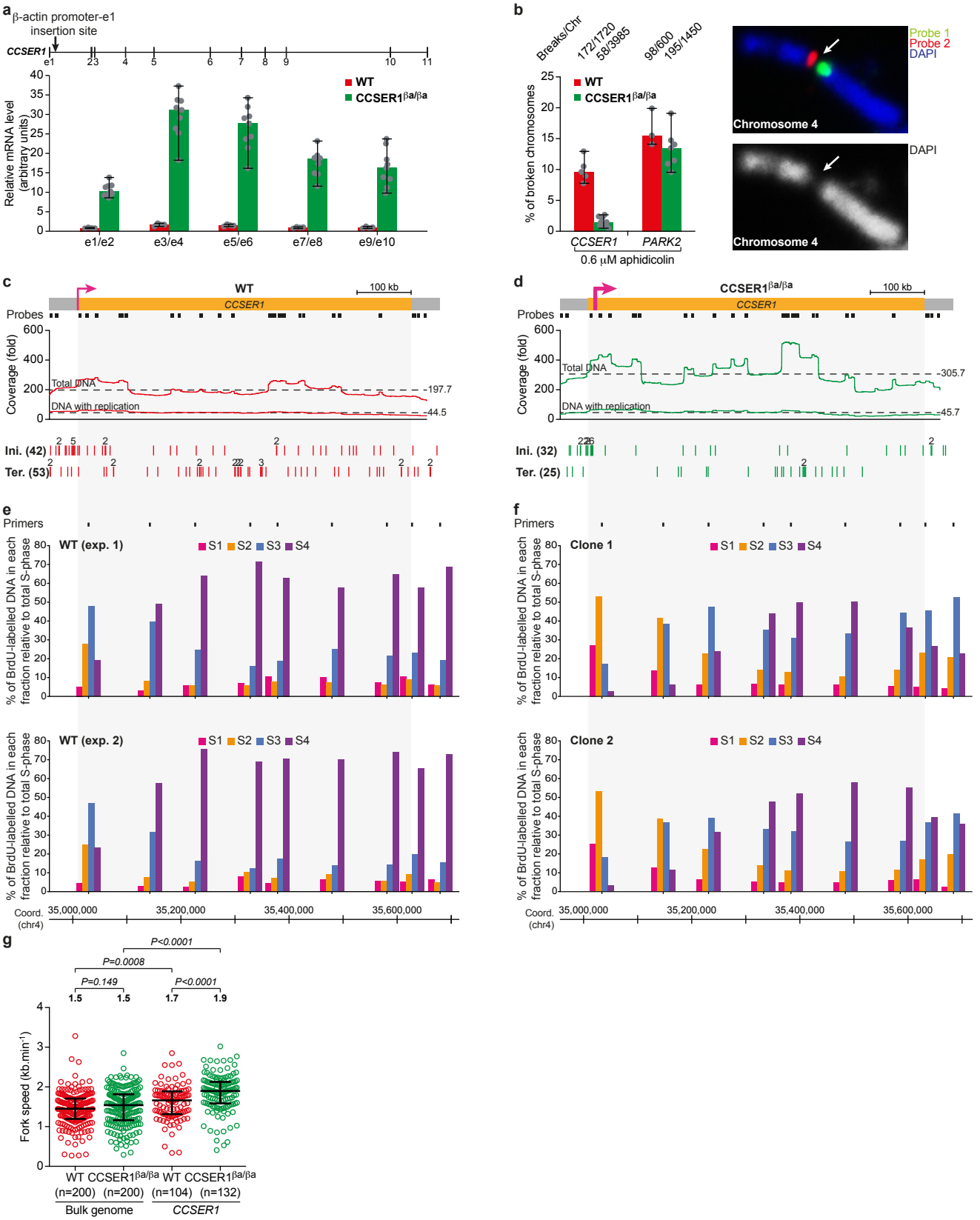
a
β-actin promoter-e1
insertion site
CCSER1
e1
23
4
5
6
7
8
9
10
11
Relative mRNA level (arbitrary units)
WT
CCSER1βa/βa
e1/e2
e3/e4
e5/e6
e7/e8
e9/e10
b
Breaks/Chr
172/1720
58/3985
98/600
195/1450
% of broken chromosomes
WT
CCSER1βa/βa
CCSER1
PARK2
0.6 μM aphidicolin
Probe 1
Probe 2
DAPI
Chromosome 4
DAPI
Chromosome 4
c
WT
100 kb
CCSER1
Probes
Coverage (fold)
Total DNA
197.7
DNA with replication
44.5
Ini. (42)
Ter. (53)
d
CCSER1βa/βa
100 kb
CCSER1
Probes
Coverage (fold)
Total DNA
305.7
DNA with replication
45.7
Ini. (32)
Ter. (25)
e
Primers
WT (exp. 1)
S1 S2 S3 S4
% of BrdU-labelled DNA in each fraction relative to total S-phase
WT (exp. 2)
S1 S2 S3 S4
% of BrdU-labelled DNA in each fraction relative to total S-phase
Coord. (chr4)
35,000,000
35,200,000
35,400,000
35,600,000
f
Primers
Clone 1
S1 S2 S3 S4
% of BrdU-labelled DNA in each fraction relative to total S-phase
Clone 2
S1 S2 S3 S4
% of BrdU-labelled DNA in each fraction relative to total S-phase
Coord. (chr4)
35,000,000
35,200,000
35,400,000
35,600,000
g
P<0.0001
P=0.0008
P=0.149
P<0.0001
1.5
1.5
1.7
1.9
Fork speed (kb.min-1)
WT
CCSER1βa/βa
WT
CCSER1βa/βa
(n=200)
(n=200)
(n=104)
(n=132)
Bulk genome
CCSER1

Figure 2. Impact of *CCSER1* transcription upregulation on its replication and fragility.

a, *CCSER1* mRNA levels relative to β-actin mRNA in WT and CCSER1$^{\beta a/\beta a}$ cells (median, extreme values and individual data points). Tested exonic junctions are indicated. Top: Map of *CCSER1* with its exons and the position of the chicken β-actin promoter-*CCSER1* exon 1 insertion site.

b, Left panel: Aphidicolin-induced breaks at *CCSER1* and *PARK2* in WT and CCSER1$^{\beta a/\beta a}$ cells (median, extreme values and individual data points; aggregate numbers are presented on top). Breaks at *PARK2* were used as a control. Right panel: Example of two-color FISH with probes flanking *CCSER1*. See Fig. 1b for details. We also verified that *CCSER1* is not fragile in the absence of aphidicolin in WT and CCSER1$^{\beta a/\beta a}$ cells (see Supplementary Fig. 4g).

c, d, Mapping of initiation and termination events along the *CCSER1* locus in WT (**c**) and CCSER1$^{\beta a/\beta a}$ (**d**) cells. From top to bottom: *CCSER1* locus (yellow box; a thin and a thick pink arrow represents active *CCSER1* promoter in WT cells and active β-actin promoter in CCSER1$^{\beta a/\beta a}$ cells, respectively) with the name of the cell line indicated above; Morse-code probes (black boxes) used in DNA-combing experiments to identify *CCSER1*; Coverage of the locus; Map of initiation and termination events. See Fig. 1c for details.

e, f, Replication timing profile of *CCSER1* in WT (**e**) and CCSER1$^{\beta a/\beta a}$ (**f**) cells. Two experiments are shown; for CCSER1$^{\beta a/\beta a}$ cells, replication timing analyses performed on two different clonal cell lines are presented. See Fig. 1e for details.

g, Replication fork progression in WT and CCSER1$^{\beta a/\beta a}$ cells in the bulk genome or in *CCSER1*. See Fig. 1g for details.

Figure 3

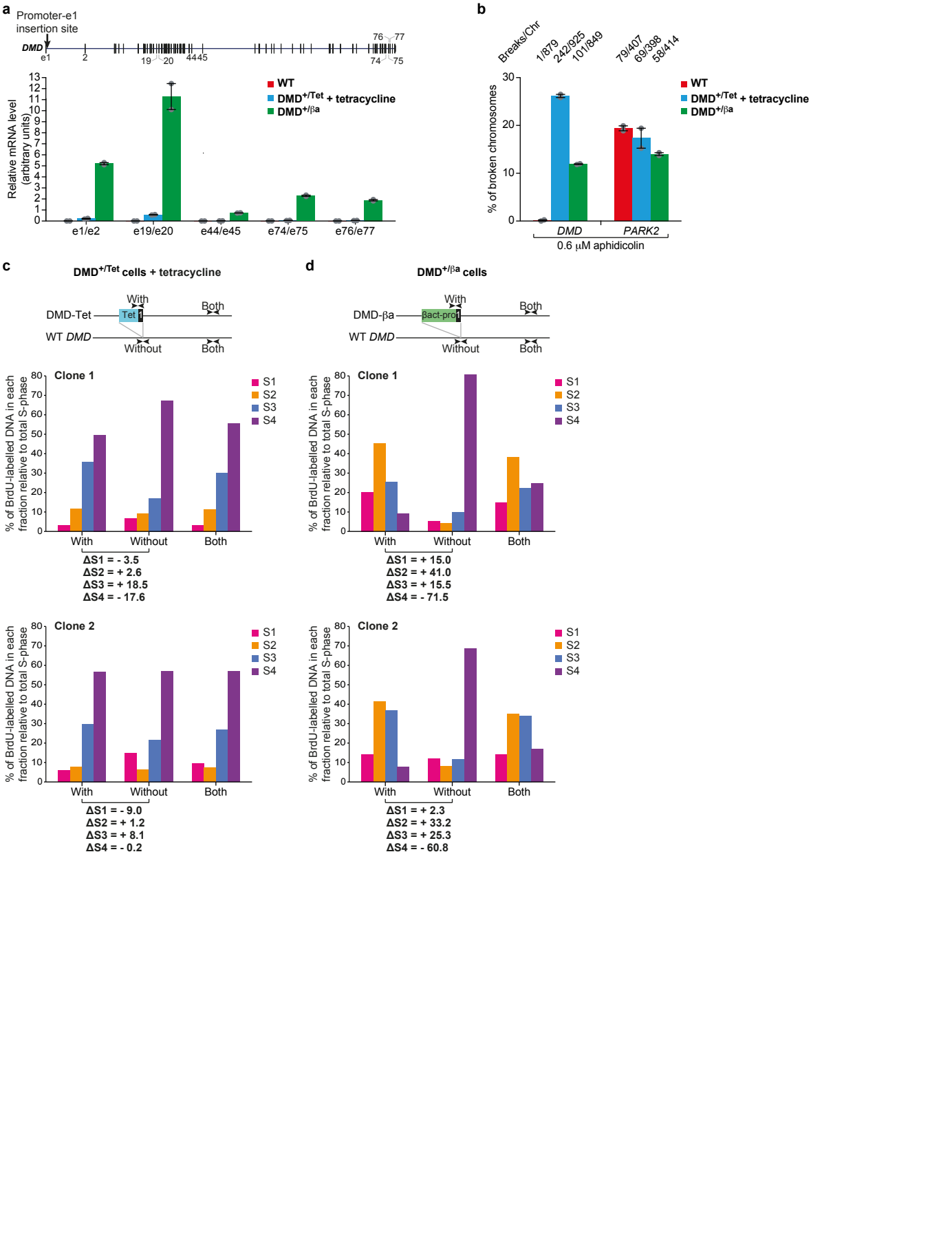
a
Promoter-e1 insertion site
DMD
e1
2
19
20
44
45
74
75
76
77
Relative mRNA level (arbitrary units)
WT
DMD+/Tet + tetracycline
DMD+/βa
e1/e2
e19/e20
e44/e45
e74/e75
e76/e77
b
Breaks/Chr
1/879
242/925
101/849
79/407
69/398
58/414
% of broken chromosomes
DMD
PARK2
0.6 μM aphidicolin
c
DMD+/Tet cells + tetracycline
DMD-Tet
Tet 1
WT DMD
With
Without
Both
Clone 1
% of BrdU-labelled DNA in each fraction relative to total S-phase
S1
S2
S3
S4
ΔS1 = - 3.5
ΔS2 = + 2.6
ΔS3 = + 18.5
ΔS4 = - 17.6
Clone 2
ΔS1 = - 9.0
ΔS2 = + 1.2
ΔS3 = + 8.1
ΔS4 = - 0.2
d
DMD+/βa cells
DMD-βa
βact-pro1
ΔS1 = + 15.0
ΔS2 = + 41.0
ΔS3 = + 15.5
ΔS4 = - 71.5
ΔS1 = + 2.3
ΔS2 = + 33.2
ΔS3 = + 25.3
ΔS4 = - 60.8

Figure 3. Impact of *DMD* transcription modulation on its replication timing and fragility.

a, *DMD* mRNA levels relative to β-actin mRNA in WT cells, $DMD^{+/Tet}$ cells with tetracycline and $DMD^{+/\beta a}$ cells (median, extreme values and individual data points). Tested exonic junctions are indicated. Top: Map of *DMD* with its exons and the position of the Tet- or β-actin promoter-*DMD* exon 1 insertion site.

b, Aphidicolin-induced breaks at *DMD* and *PARK2* in WT cells, $DMD^{+/Tet}$ cells with tetracycline and $DMD^{+/\beta a}$ cells (median, extreme values and individual data points; aggregate numbers are presented on top). Breaks at *PARK2* were used as a control.

c, d, Replication timing at the Tet- (**c**) and β-actin (**d**) promoter insertion site. BrdU pulse-labelled $DMD^{+/Tet}$ cells with tetracycline and $DMD^{+/\beta a}$ cells were sorted into four S-phase fractions, and neo-synthesized DNA was quantified by real-time PCR using primer pairs specific of the DMD-Tet or DMD-βa allele (with), of the WT *DMD* allele (without) or hybridizing on the two alleles (both). The differences in replication timing at the target site following promoter integration ($\Delta S = S_{with} - S_{without}$) were calculated for each sub-fraction. Results for two different clonal cell lines are presented for each construct.

Figure 4

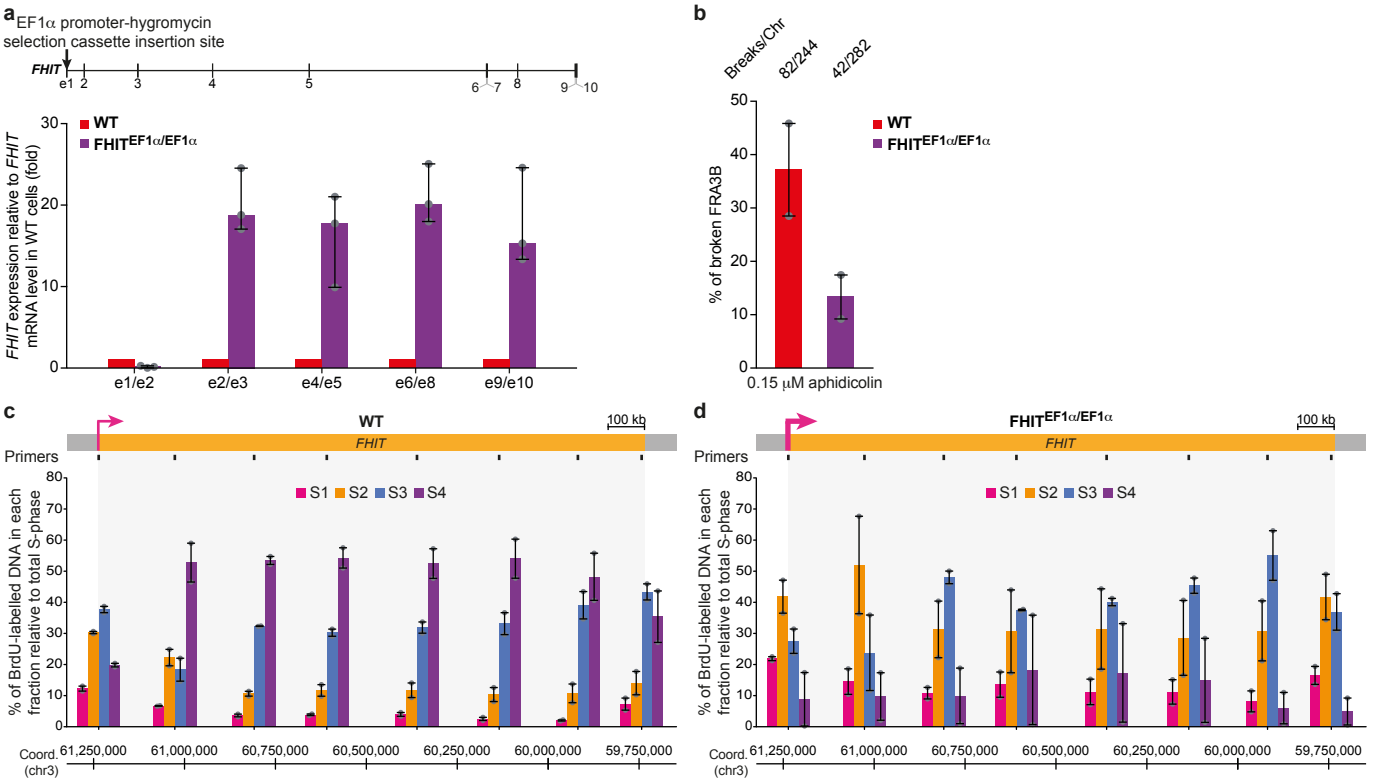
a
EF1α promoter-hygromycin selection cassette insertion site
FHIT
e1 2 3 4 5 6 7 8 9 10
WT
FHITEF1α/EF1α
FHIT expression relative to FHIT mRNA level in WT cells (fold)
e1/e2
e2/e3
e4/e5
e6/e8
e9/e10
b
Breaks/Chr
82/244
42/282
WT
FHITEF1α/EF1α
% of broken FRA3B
0.15 μM aphidicolin
c
WT
FHIT
100 kb
Primers
S1 S2 S3 S4
% of BrdU-labelled DNA in each fraction relative to total S-phase
Coord. (chr3)
61,250,000
61,000,000
60,750,000
60,500,000
60,250,000
60,000,000
59,750,000
d
FHITEF1α/EF1α
FHIT
100 kb
Primers
S1 S2 S3 S4
% of BrdU-labelled DNA in each fraction relative to total S-phase
Coord. (chr3)
61,250,000
61,000,000
60,750,000
60,500,000
60,250,000
60,000,000
59,750,000

Figure 4. Impact of *FHIT* transcription upregulation on its replication timing and fragility in human HCT116 cells.

a, *FHIT* expression in WT and FHIT$^{EF1\alpha/EF1\alpha}$ cells relative to *FHIT* mRNA in WT cells (median, extreme values and individual data points). *FHIT* mRNA level was normalized to 3 housekeeping genes (*POLR2F*, *RPL11* and *PPIB*). Tested exonic junctions are indicated. Top: Map of *FHIT* with its exons and the position of the EF1α promoter-hygromycin selection cassette insertion site.

b, Aphidicolin-induced breaks at *FHIT* in WT and FHIT$^{EF1\alpha/EF1\alpha}$ cells (median, extreme values and individual data points; aggregate numbers are presented on top).

c, d, Replication timing profile of *FHIT* in WT (**c**) and FHIT$^{EF1\alpha/EF1\alpha}$ (**d**) cells (median and extreme values of 2 experiments). See Fig. 1e for details.