Cell Reports, Volume *32*

Supplemental Information

Replication Stress Induces Global Chromosome Breakage in the Fragile X Genome

Arijita Chakraborty, Piroon Jenjaroenpun, Jing Li, Sami El Hilali, Andrew McCulley, Brian Haarer, Elizabeth A. Hoffman, Aimee Belak, Audrey Thorland, Heidi Hehnly, Carl Schildkraut, Chun-long Chen, Vladimir A. Kuznetsov, and Wenyi Feng

Supplemental Figure S1

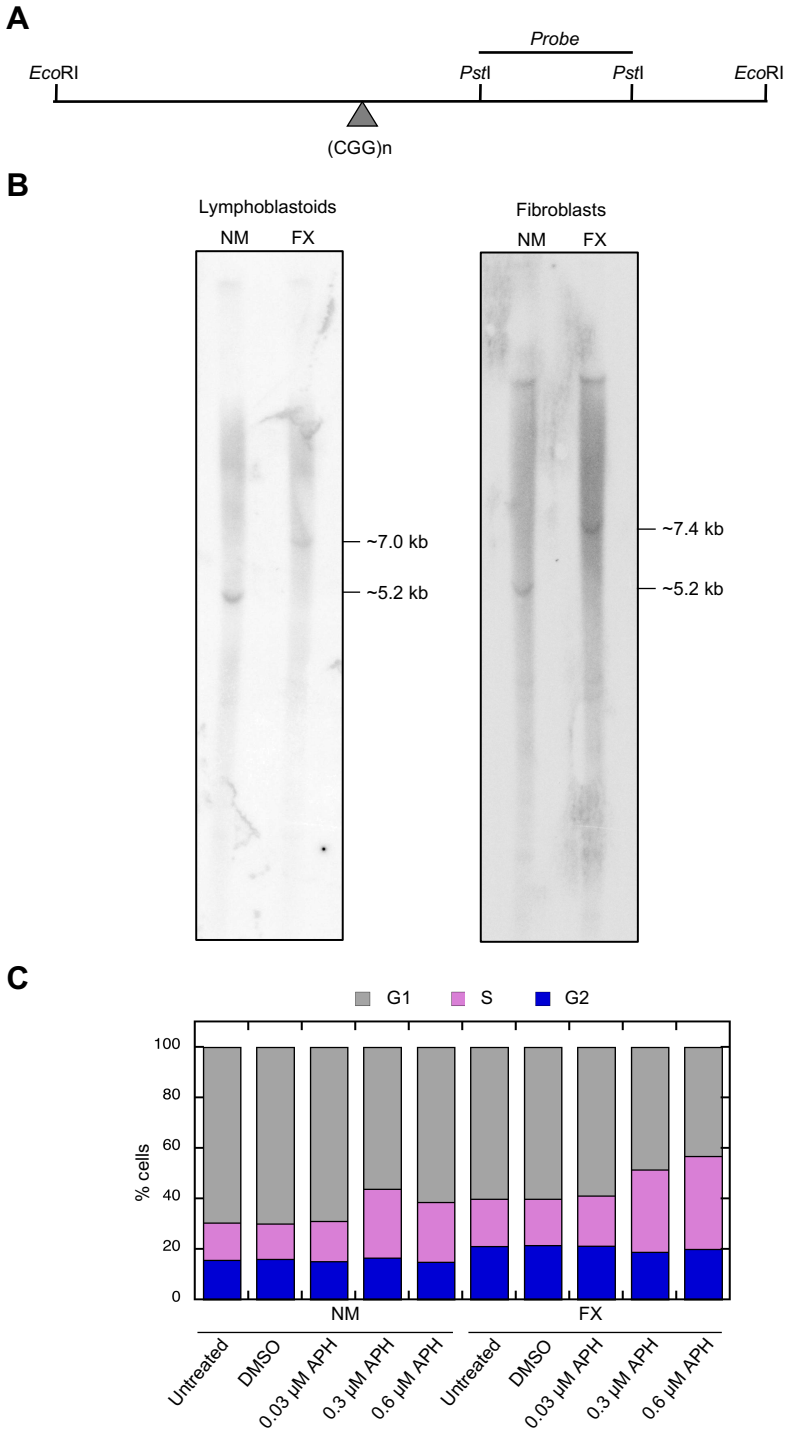

Figure S1. Related to Figure 1A. Validation of CGG repeat expansion at *FMR1* locus and the absence of FMRP expression in the Fragile X cell lines used in this study. (A) The restriction map of the 5.2 kb *Eco*RI fragment in the *FMR1* locus and probe position used for Southern blots are shown. The probe was synthesized by PCR amplification from genomic DNA isolated from GM06990 using the following primers: FMR1-F, 5'-TGGCTTCTCTTTTCCGGTCT-3'; FMR1-R, 5'-GGGTTACCTTTTGCCTCCTT -3'. (**B**) Southern blots of *Eco*RI-digested genomic DNA from the indicated cell lines. A band corresponding to 5.2 kb (containing 30-50 CGG repeats) was seen in the normal cell lines. In contrast, two bands corresponding to ~7.0 kb and ~7.4 kb were identified in the FX lymphoblastoid and fibroblast lines, which correspond to ~590 and 730 CGG repeat expansion, respectively. **(C)** Cell cycle distribution analyzed by flow cytometry. Three independent experiments were conducted, and one representative experiment is shown.

Supplemental Figure S2

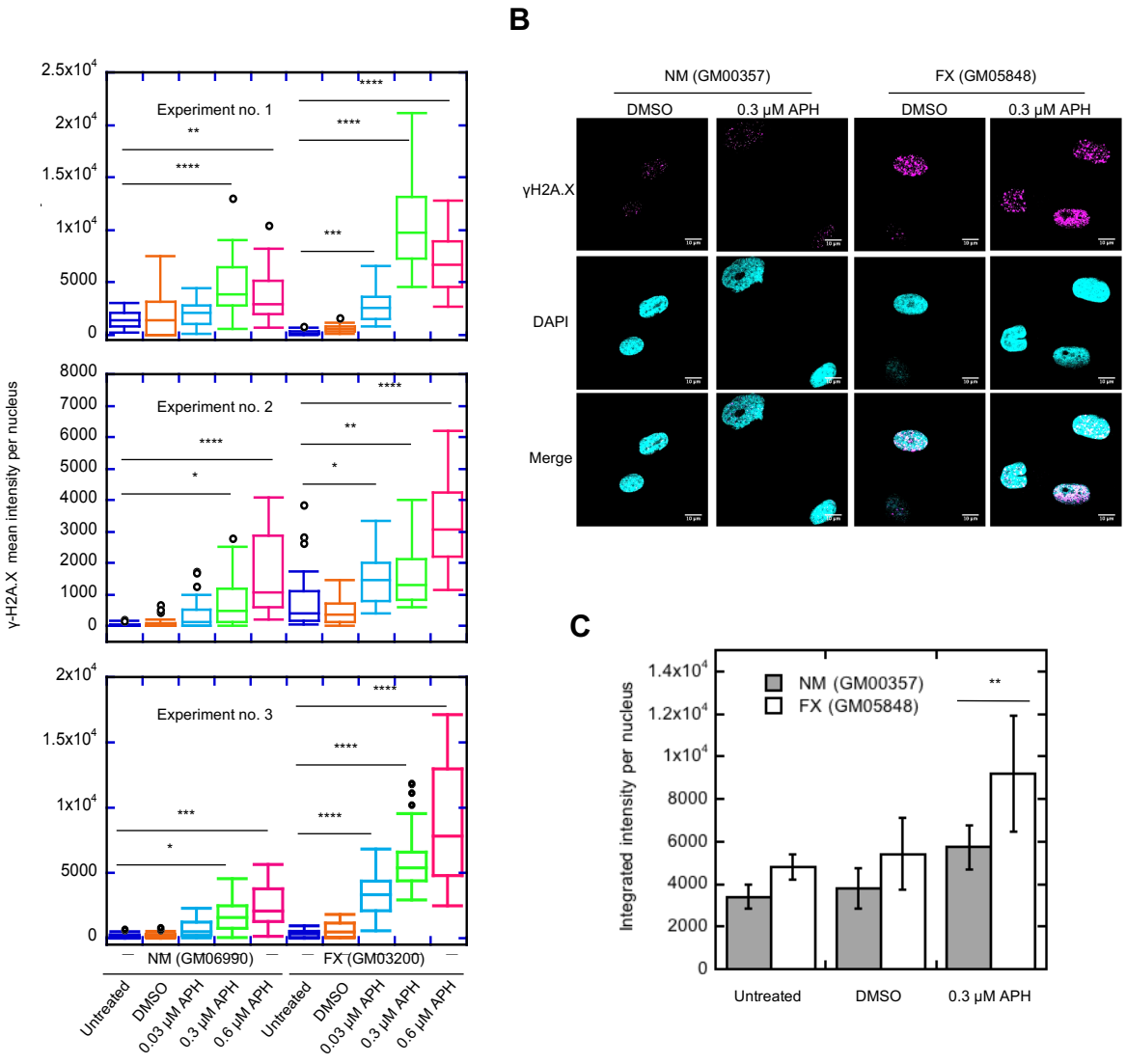

Figure S2. Related to Figure 1B. Lymphoblastoid cells and fibroblasts from FX have increased DNA damage. (A) Box plots for γ-H2A.X immunofluorescence staining of lymphoblastoid cells under the indicated conditions from **Figure 1B**. Plotted are mean intensity values of fluorescence per nucleus in three independent experiments (n≥28 nuclei counted for each experiment). One-way ANOVA test followed by Tukey's multiple comparison test were performed. **(B&C)** Fibroblast cell line from a FX individual shows increased DNA damage compared to a sex and age matched control fibroblast cell line. Two independent experiments were performed, and one representative experiment is shown here. At least 41 nuclei per sample were analyzed in each experiment. Representative images for γH2A.X staining in FX and NM cells under the indicated treatment (B) and quantification of γH2A.X signals per nucleus (C) are shown. Error bars indicate standard deviation. Two-way ANOVA test followed by Sidak's multiple testing was performed. *p<0.05, **p<0.01, ***p<0.001 and ****p<0.0001. Scale bar, 10 μm.

Supplemental Figure S3

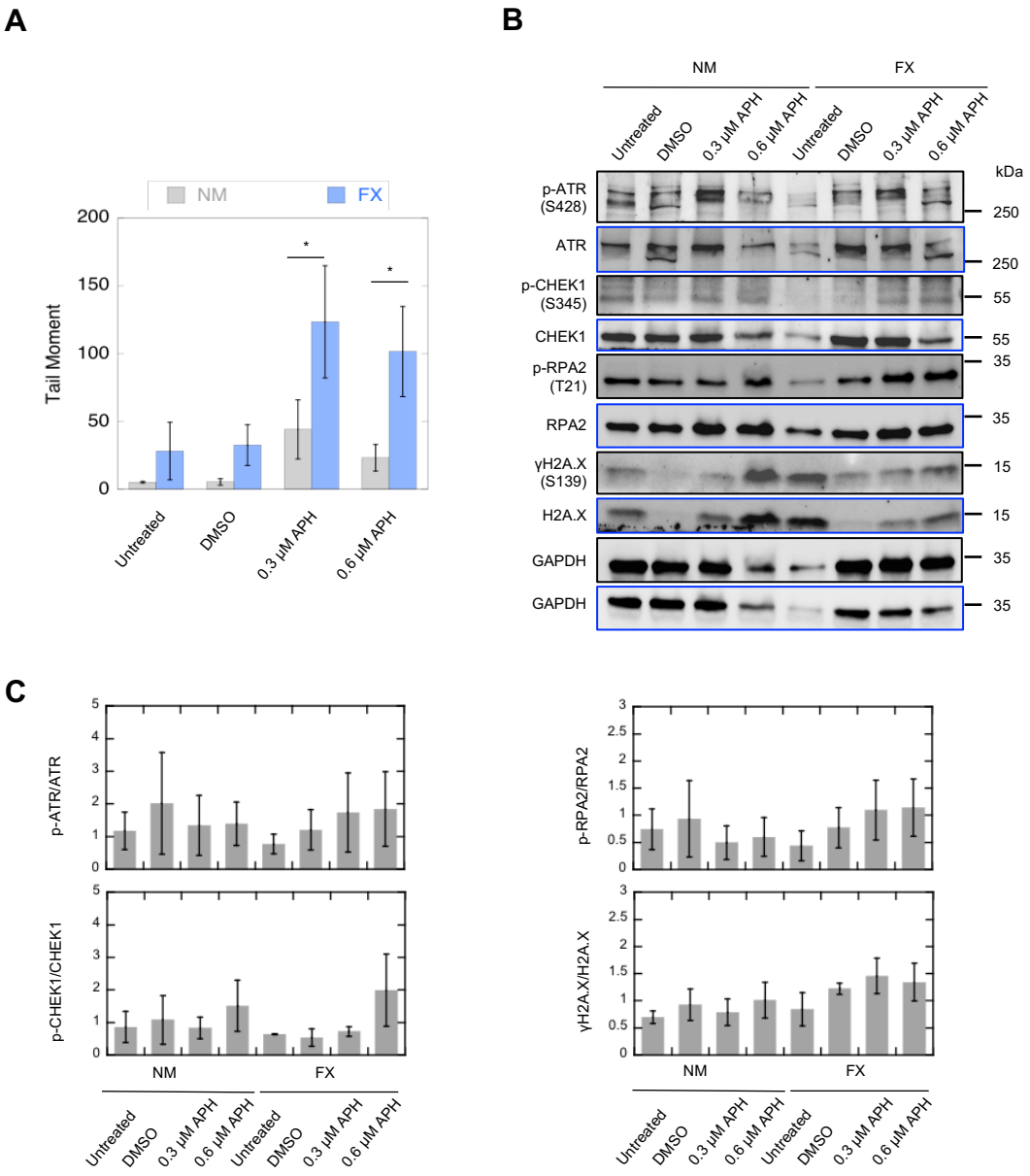

Figure S3. Related to Figures 1E&F. DNA damage is elevated and DDR is not impaired in FX cells. (**A**) DSB formation measured by neutral Comet assay. Three independent experiments were performed (n>50 in each) and all trended similarly. A representative experiment is shown for Tail Moment measurement. One-way ANOVA followed by Sidak's multiple testing was performed. Error bars denote standard errors of mean and *p<0.05. (**B**) Whole cell lysates from NM and FX lymphoblastoid cells western blotted for proteins in the DDR pathway. Blots probed with antibodies against phospho(p)-proteins and total proteins are shown in black and blue bordered panels, respectively. Both sets of blots were controlled for loading using GAPDH. (**C**) Quantification of phosphorylation of DDR proteins in (A). Error bars indicate standard error of mean.

Supplemental Figure S4

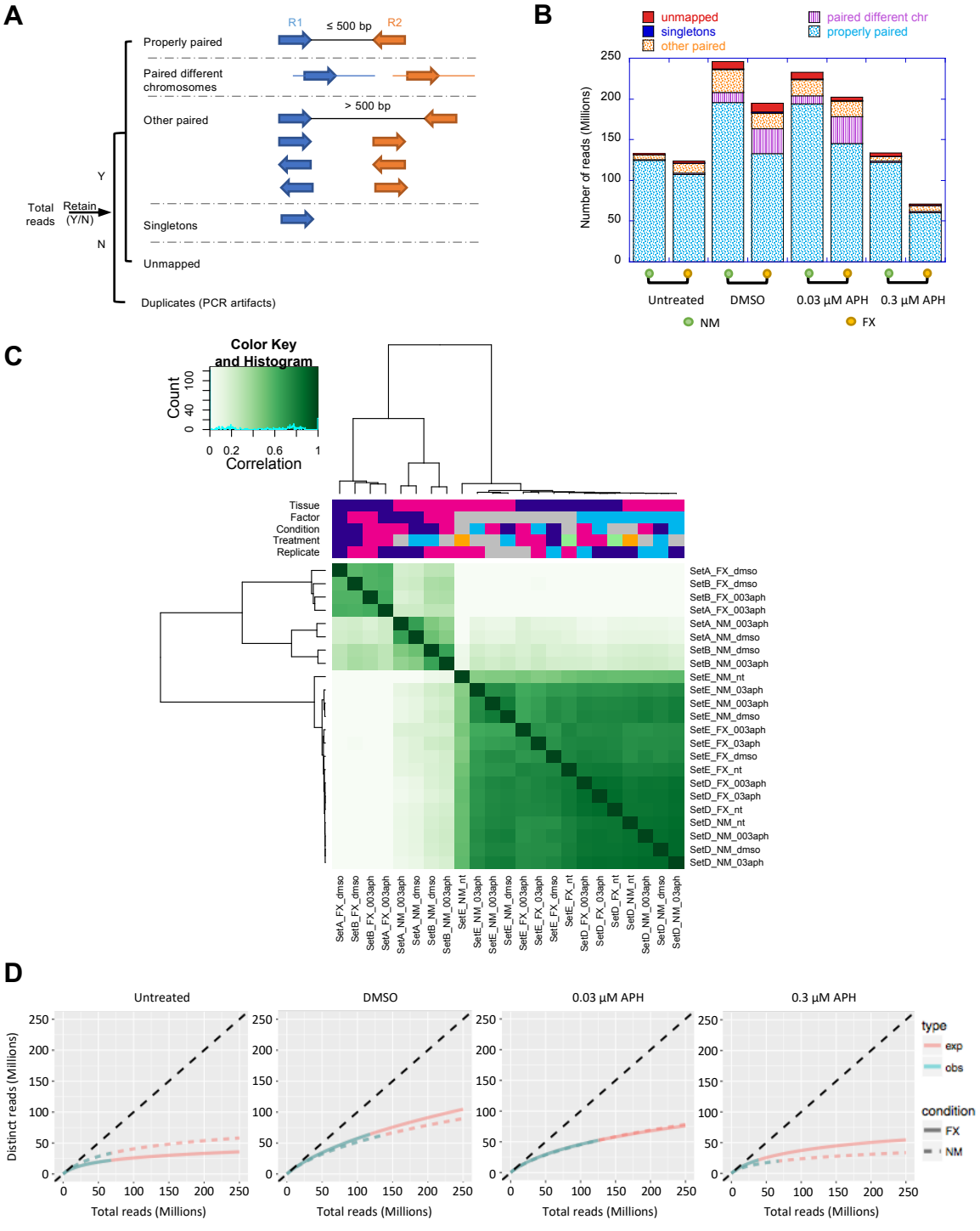

Figure S4. Related to Figure 3. Break-seq data quality check. (A&B) Break-seq library quality assessed by read classifications. "R1" and "R2", Read 1 and Read 2, respectively. Samples with the same treatment were merged to assess overall sequencing depth in (B). **(C)** Break-seq experiments correlation heatmap generated by plot (FX.breakseq) in DiffBind. Note that FX_dmso has three instead of four replicates due to the removal of a dubious experiment (SetD_FX_dmso). **(D)** Break-seq library complexity of FX and NM cells ("obs" and "exp" stand for observed and expected library complexities, respectively). Libraries from biological replicates for each treatment condition were merged for complexity measure.

Supplemental Figure S5

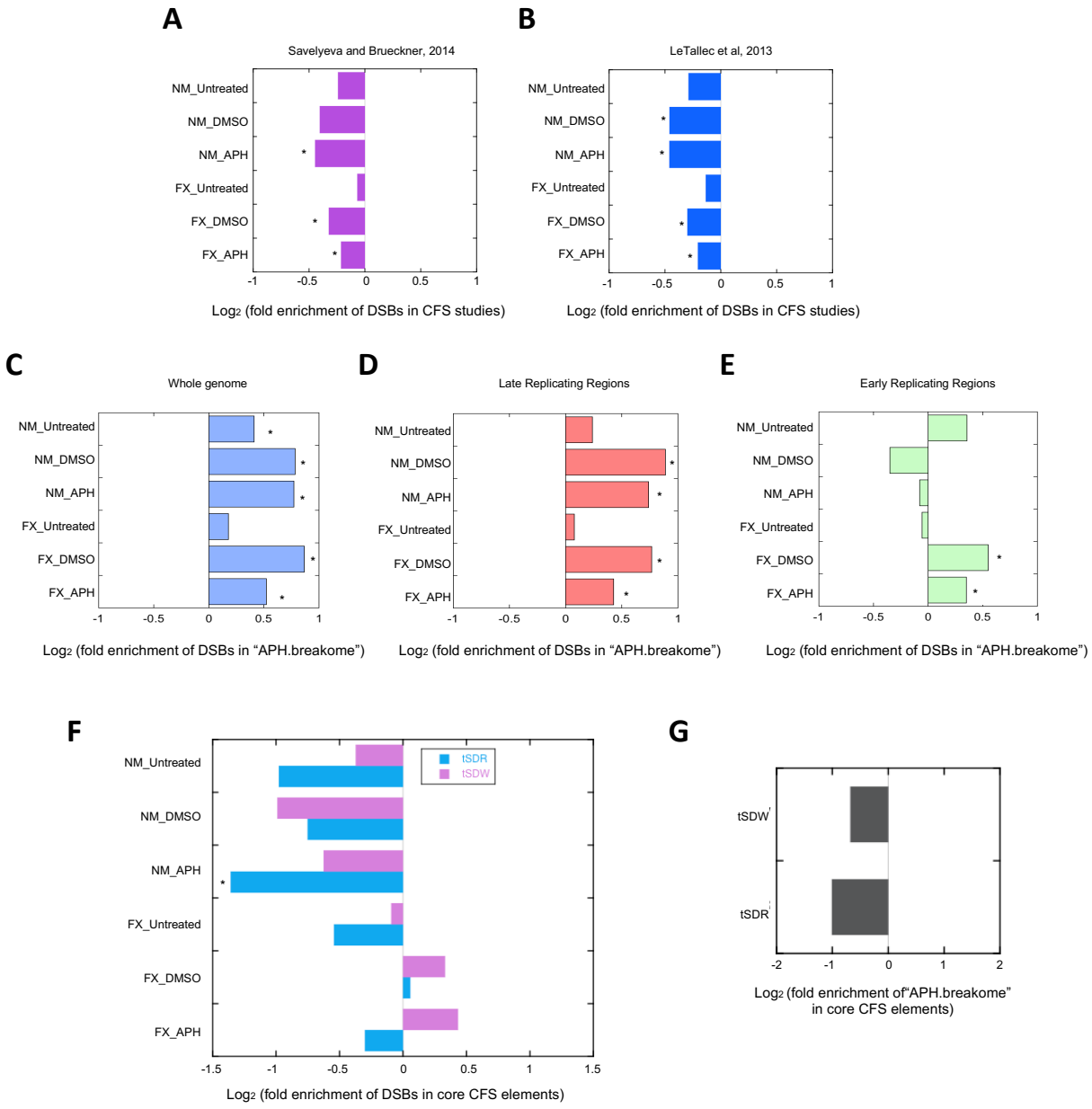

Figure S5. Related to Figure 4D. Genomic Association Test (GAT) for correlations between DSBs in our study and CFSs reported in literature. (**A-B**) Correlation between DSBs and reported CFSs. Log2 transformation of the fold enrichment values (ratios of observed to expected number of DSBs) are reported for CFSs mapped by Salvelyeva and Bruekner, 2014 (**A**) and LeTallec et al, 2013 (**B**). Correlation between those DSBs that are concordant with the "APH.breakome" in the whole genome (**C**), the late replicating regions (**D**), and the early replicating regions (**E**). (**F**) Correlation between DSBs and the late-replicating CFS core elements containing large transcribed genes, defined as tSDWs (transcribed Significantly Delayed Windows) and tSDRs (transcribed Significantly Delayed Regions), in Brison et al. 2019. (**G**) Correlation between "APH.breakome" and the late-replicating CFS core elements containing large transcribed genes as in (F). Those samples marked with an asterisk indicate p values ≤ 0.001.

Supplemental Figure S6

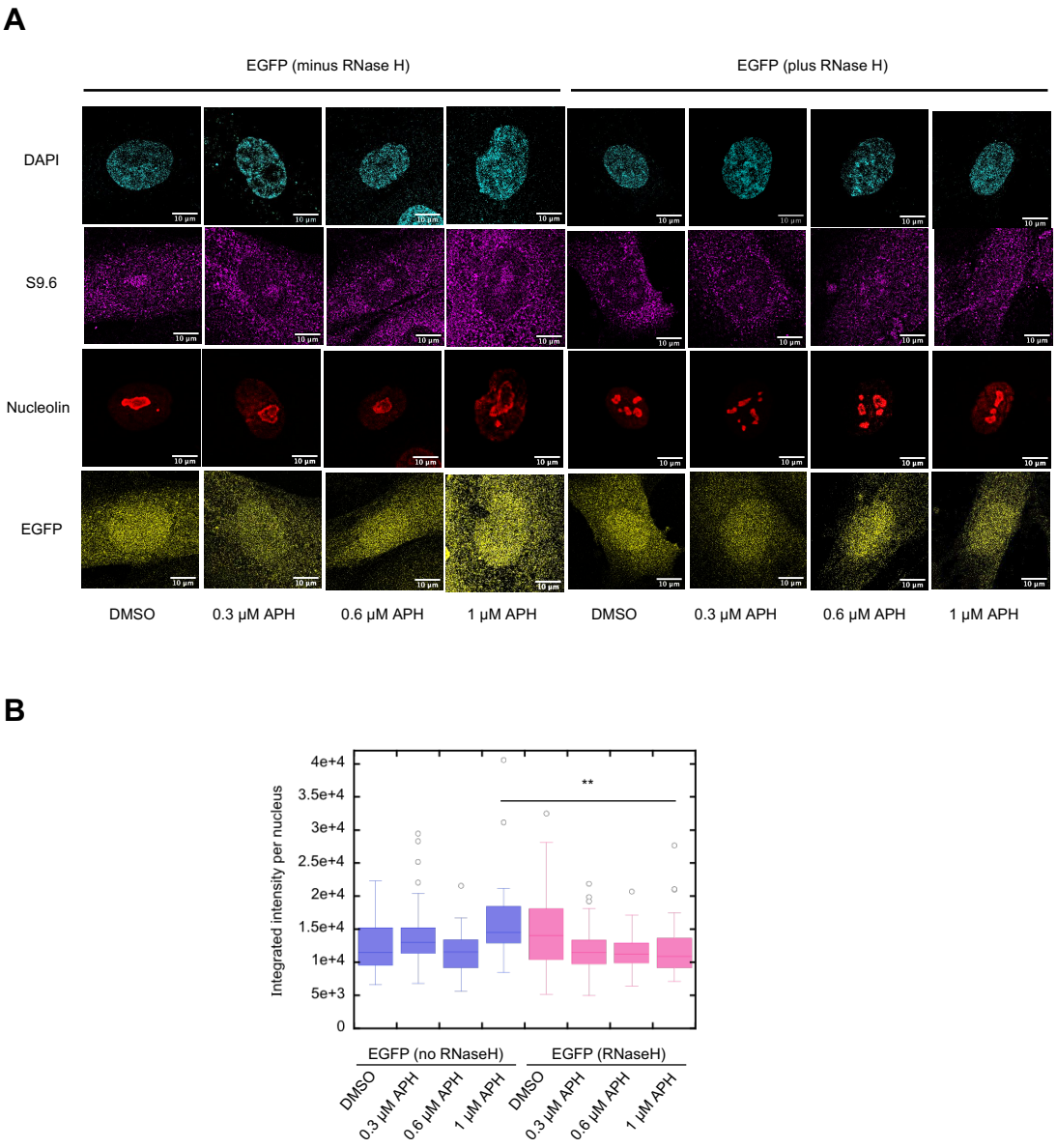

Figure S6. Related to Figure 5. Nuclear RNA:DNA hybrid signals detected by S9.6 antibody is labile to RNase H treatment. (**A**) Representative confocal images of immunofluorescence staining with S9.6 antibody in FX cells expression EGFP untreated or treated with RNaseH prior to staining. Scale bar, 10 μm. (**B**) Quantification of nuclear S9.6 signals in three independent experiments (n≥36 in each) using one-way ANOVA with multiple comparisons where **p<0.01.

Supplemental Figure S7

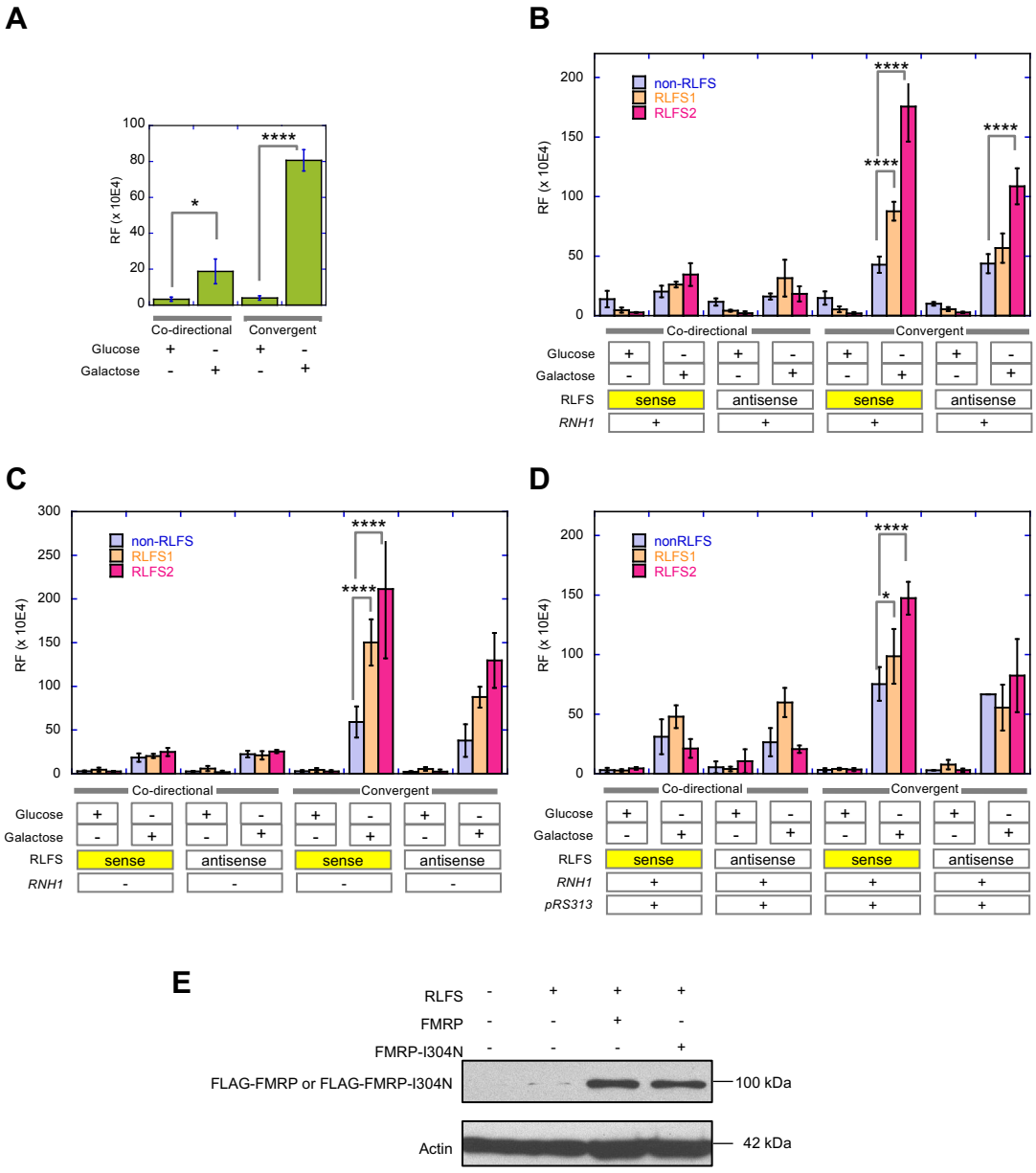

Figure S7. Related to Figure 6. Control experiments for RLFS-induced DNA breakage and recombination frequency (RF) in yeast. (**A**) RF is dependent on transcriptional activation. One-way ANOVA followed by Tukey's multiple testing was performed. (**B**) Insertions of RLFS elements from the human genome can further induce RF, specifically when inserted in the "sense" orientation with respect to transcription. (**C**) The RLFS-induced RF can be further enhanced by deletion of *rnh1*, the gene encoding for RNase H1. Note the change of scale on the Y-axis in (C). (**D**) Cells bearing a second plasmid, pRS313, which appeared to have little effect on RF, serve as control for the experiments in Fig. 3B. From (A) to (D) *p<0.05 and ****p<0.0001. **E**) FMRP and FMRP-I304N show similar expression levels in yeast. Cells transformed with a plasmid with or without an RLFS in the LEU2 gene cassette, together with a plasmid with or without the CMV-driven and FLAG-tagged FMR1 or FMR1-I304N, were analyzed by Western blot using the anti-FLAG antibody. Actin served as a loading control.

SUPPLEMENTAL TABLES

Table S1. Related to figure 4C. Number of consensus DSB peaks (detected in at least 2 replicates) in all categories defined by strain/treatment/comparison combinations.

DSB category[1]	Number of consensus DSBs
FX.APH	18473
FX.APH003	16796
FX.APH03	2112
FX.DMSO	9167
FX.NT	4149
FXdmso.FXaph.overlap	6177
FXdmso.FXaph.uniquetoFXaph	12296
FXdmso.FXaph.uniquetoFXdmso	2984
FXdmso.FXnt.overlap	322
FXdmso.FXnt.uniquetoFXdmso	8845
FXdmso.FXnt.uniquetoFXnt	3827
NM.APH	7002
NM.APH003	3506
NM.APH03	2161
NM.DMSO	3927
NM.NT	2111
NMdmso.NMaph.overlap	3209
NMdmso.NMaph.uniquetoNMaph	3792
NMdmso.NMaph.uniquetoNMdmso	714
NMdmso.NMnt.overlap	1651
NMdmso.NMnt.uniquetoNMdmso	2276
NMdmso.NMnt.uniquetoNMnt	458
FXdmso.NMdmso.overlap	644
FXdmso.NMdmso.uniquetoFXdmso	8523
FXdmso.NMdmso.uniquetoNMdmso	3283
FXnt.NMnt.overlap	1753
FXnt.NMnt.uniquetoFXnt	2396
FXnt.NMnt.uniquetoNMnt	358
FXaph.NMaph.overlap	4369
FXaph.NMaph.uniquetoFXaph	14104
FXaph.NMaph.uniquetoNMaph	2633

[1]Strain=FX, NM; Treatment=NT (untreated), DMSO, APH003 (0.03 μM APH), APH03 (0.3 μM APH), APH (composite of APH003 and APH03); Comparison=overlap (shared by two samples), uniquetoXXxx (unique to the sample indicated).

Table S2. Related to figure 5G-H. Select top Gene Ontology terms ($p < 0.001$) for "Biological pathways" associated with DSBs in the indicated groups.

DSB Group	Biological pathway	P value	# of genes
FX_UNTREATED	neuron projection development	1.12E-05	100
	nervous system development	2.08E-05	214
	regulation of cell morphogenesis	2.08E-05	66
	synapse organization	5.29E-05	40
	regulation of neuron projection development	6.25E-05	58
	neuron development	8.09E-05	109
	neuron cell-cell adhesion	0.00015	9
	neuron projection morphogenesis	0.00018	70
	ion transmembrane transport	0.00025	107
	inorganic ion transmembrane transport	0.00048	85
FX_DMSO	cell projection organization	0.00030	187
	localization	0.00030	717
	cell part morphogenesis	0.00050	131
	inorganic ion transmembrane transport	0.00052	118
	neuron development	0.00075	146
	cell projection morphogenesis	0.00086	126
FX_APH	positive regulation of GTPase activity	6.76E-06	180
	regulation of cell morphogenesis	7.35E-06	139
	nervous system development	1.15E-05	504
	regulation of cell projection organization	4.09E-05	152
	neuron projection development	5.56E-05	213
	regulation of neuron projection development	9.09E-05	119
	cell adhesion	0.000107	405
	chemical synaptic transmission	0.000107	163
	movement of cell or subcellular component	0.000199	415
	cell projection assembly	0.000212	113
	vesicle-mediated transport	0.000347	343
	synaptic vesicle localization	0.000569	46
	microtubule-based process	0.000596	160
	synaptic vesicle cycle	0.000681	41
	dendrite development	0.000733	63
	cytoskeleton organization	0.000784	271
NM_UNTREATED	None		
NM_DMSO	None		
NM_APH	regulation of cell projection organization	2.20E-06	108
	regulation of cell morphogenesis	1.56E-05	94
	cell projection organization	3.97E-05	201
	regulation of neuron projection development	0.000217	80
	neuron projection development	0.000559	135
	neuron development	0.000559	154
	nervous system development	0.000646	305
	cellular component morphogenesis	0.000646	186
FX_DMSO_RLFS_overlap	flavonoid glucuronidation	0.000136	8
	cellular glucuronidation	0.000208	8
	uronic acid metabolic process	0.000499	8
	flavonoid metabolic process	0.000499	8
	flavonoid biosynthetic process	0.000499	7
	glucuronate metabolic process	0.000499	8
FX_APH_RLFS_overlap	single-organism membrane organization	0.001285	84

Table S3. Related to figure 6A. Correlation between DSBs in various groups and 169222 computationally predicted RLFSs (Wongsurawat et al., 2012) and 108011 composite DRIP-seq signals merged from all DRIP-seq data sets (NT2 and K562) (Sanz et al., 2016). Note the merged DRIP-seq dataset has 6.6 times the coverage of the RLFSs.

DSB group	Number of DSBs	Query (RLFS or DRIP-seq signals)	Number of DSBs overlapped with RLFS or DRIP-seq signals	P value Left	P value Right	P value Two-tail	Enrich-ment Ratio
FX.NT	4149	RLFS	133	2.34E-43	1	4.37E-43	0.358
FX.DMSO	9167	RLFS	1294	1	3.71E-37	6.24E-37	1.493
FX.APH	18473	RLFS	2323	1	1.25E-66	2.02E-66	1.498
NM.NT	2111	RLFS	129	7.31E-13	1	1.36E-12	0.552
NM.DMSO	3927	RLFS	659	1	1.22E-34	1.82E-34	1.746
NM.APH	7002	RLFS	554	0.045261	0.95877	0.089567	0.927
FX.NT	4149	DRIP-seq	526	7.62E-14	1	1.54E-13	0.717
FX.DMSO	9167	DRIP-seq	2658	1	1.07E-155	1.55E-	1.914
FX.APH	18473	DRIP-seq	4137	1	1.36E-80	2.21E-80	1.422
NM.NT	2111	DRIP-seq	389	0.64755	0.37359	0.73422	1.02
NM.DMSO	3927	DRIP-seq	1229	1	2.21E-90	2.98E-90	2.083
NM.APH	7002	DRIP-seq	1326	1	1.57E-06	2.93E-06	1.157

P values were calculated using bedtools Fisher's Exact Test. "NT", untreated.

Table S4. Related to figure 6A. Statistical analysis of absolute distance between RLFSs and DSBs in the indicated categories on each chromosome using GenometriCorr. Shown are the projection test p-values (P) and projection test observed to expected ratios (R). Samples with R values greater than 3.5 are shown in bold.

Chr	FX.Untreated		FX.DMSO		FX.APH		NM.Untreated		NM.DMSO		NM.APH	
	P	R	P	R	P	R	P	R	P	R	P	R
1	0.355	1.04	**2.22e-**	**3.79**	1.07e-	2.76	0.224	1.23	0.086	1.51	0.353	0.81
2	0.314	0.64	**2.21e-**	**3.62**	1.55e-	2.62	0.236	1.21	**1.73e-**	**3.79**	0.010	1.85
3	0.446	0.69	1.29e-	3.02	1.81e-	2.57	0.329	0.00	0.024	2.28	0.131	1.43
4	0.283	0.39	2.39e-	3.00	1.44e-	3.22	0.355	0.80	0.036	2.20	0.199	1.28
5	0.217	0.34	1.09e-	3.24	1.63e-	2.79	0.244	0.00	0.007	2.55	0.148	1.38
6	0.090	0.00	**6.36e-**	**3.74**	5.08e-	2.06	0.336	0.00	0.134	1.53	0.028	0.00
7	0.489	0.86	6.30e-	2.51	4.01e-	2.57	0.108	0.00	0.104	1.48	0.126	1.33
8	0.150	0.29	4.34e-	2.96	1.06e-	2.08	0.182	1.36	0.203	1.29	0.131	0.40
9	0.301	0.41	3.91e-	3.21	4.19e-	3.19	0.174	1.39	0.016	2.17	0.109	0.46
10	0.137	0.28	2.85e-	3.27	3.85e-	2.80	0.302	1.04	0.088	1.63	0.373	1.01
11	0.070	0.23	7.45e-	2.21	0.001	1.64	0.118	0.00	0.047	1.70	0.123	0.52
12	0.044	0.00	**4.91e-**	**4.04**	7.32e-	2.74	0.209	0.00	0.076	1.70	0.350	1.04
13	0.372	0.00	**9.55e-**	**6.96**	**8.71e-**	**5.43**	0.347	0.00	0.482	0.00	0.260	0.00
14	0.190	1.33	**8.41e-**	**3.51**	**6.05e-**	**3.81**	0.035	2.78	**5.87e-**	**4.21**	0.103	1.62
15	0.109	1.65	**3.62e-**	**4.23**	**2.64e-**	**3.92**	0.377	0.00	0.054	1.96	0.088	1.63
16	0.050	0.21	**0.00e-**	**5.66**	**0.00e-**	**4.52**	0.484	0.76	0.060	1.62	0.291	1.11
17	0.056	0.22	2.62e-	3.28	7.21e-	3.15	0.322	0.43	0.353	1.04	0.165	1.27
18	0.103	1.78	**1.14e-**	**3.96**	3.54e-	2.76	0.144	1.49	0.061	2.23	0.237	1.19
19	0.057	0.45	2.32e-	1.94	2.88e-	1.91	0.162	0.44	0.429	0.99	0.065	0.63
20	0.456	0.80	5.62e-	2.83	9.08e-	3.16	0.135	1.56	0.039	2.14	0.218	1.23
21	0.230	0.00	**5.01e-**	**7.10**	**2.22e-**	**6.00**	0.426	0.00	0.014	2.81	0.432	0.89
22	0.093	0.00	**1.80e-**	**5.34**	**3.99e-**	**3.76**	0.462	0.64	0.016	2.14	0.009	2.15
X	0.184	1.35	0.081	1.48	0.098	1.33	0.113	1.63	0.024	2.13	0.077	1.55
ALL	5.47e-	0.52	0	3.16	0	2.68	0.041	0.71	8.93e-	1.97	0.150	1.08

Table S5. Related to figure 6A. Number of genes containing DSBs in FX cells that overlap RLFSs or DRIP-seq signals.

DSB Group	Total DSBs	Genes containing DSBs*	Genes containing DSBs that overlap RLFSs*	Genes containing DSBs that overlap DRIP-seq signals*
FXdmso.FXaph.overlap	6177 (28.8%)	4121 (27.8%)	894 (33.1%)	2364 (37.8%)
FXdmso.FXaph.uniquetoFXaph	12296 (57.3%)	8715 (58.8%)	1427 (52.9%)	3113 (49.8%)
FXdmso.FXaph.uniquetoFXdmso	2984 (13.9%)	1990 (13.4%)	378 (14.0%)	778 (12.4%)
FXdmso.FXaph.total	21457 (100%)	14826 (100%)	2699 (100%)	6255 (100%)
NMdmso.NMaph.overlap	3209 (41.6%)	2484 (41.4%)	523 (50.7%)	1240 (41.3%)
NMdmso.NMaph.uniquetoNMaph	3792 (49.2%)	2524 (42.1%)	153 (14.8%)	718 (23.9%)
NMdmso.NMaph.uniquetoNMdmso	714 (9.3%)	990 (16.5%)	356 (34.5%)	1042 (34.7%)
NMdmso.NMaph.total	7715 (100%)	5998 (100%)	1032 (100%)	3000 (100%)

* Percentages in parentheses indicate the percentage of DSBs or genes in each of the three categories (e.g., XX.XX.overlap, XX.XX.uniquetoXXaph, and XX.XX.uniquetoXXdmso) over the sum of the three categories

Table S6. Related to Key resources table for oligonucleotides. Oligonucleotides used in this study is listed below.

Name	Source
5'-CGAGGTTAACATGGTGAGCAAGGGCGAGGAG-3' (pcDNA3_RMEGFP_JL_FWD2)	This paper
5'-ATTCGTTAACCTTGTACAGCTCGTCCATGCC -3' (pcDNA3_RMEGFP_JL_REV2)	This paper
5'-TAACGAATTCATGGACTACAAGGACGACGAT-3' (pFRT_RMFMRP_JL_FWD2)	This paper
5'-GGTAGAATTCTTAGGGTACTCCATTCACGAG-3' (pFRT_RMFMRP_JL_REV2)	This paper
5'-TCagatctTCAGGCTGCACATTCTTTTC-3' (control-1_F)	This paper
5'-CTCggatccTGCTTTCACTGCAGTTCC-3' (control-1_R)	This paper
5'-CTCagatctTGATAATTACAAGGTACACGTTATTGC-3' (control-2_F)	This paper
5'-CTCggatccTTGGTTAGGATAATAAGCACTATGG-3' (control-2_R)	This paper
5'-CTCagatctGTAGACGCCTCACCTTCTGC-3' (RLFS-1_F)	This paper
5'- CTCggatccTGCGGGTGTAAACACTGAAA-3' (RLFS-1_R)	This paper
5'- CTCagatctCATAACTAAGCACTGTATGCC-3' (RLFS-2_F)	This paper
5'- CTCggatccCCTAGGGACAAGGGGAGGTA-3' (RLFS-2_R)	This paper
5'-CTGTGTACATGAGCAAAGACAAAGATATC-3' (SHE2-Y-F)	This paper
5'-CTGGAATTCTCAGTTTTTCAATTTACC -3' (SHE2-Y-R)	This paper
5'-CTGGGAATCCCCTAATTATGGCAAGGCAAG-3' (RNH1-Y-F)	This paper
5'-CTGGCGGCCGCACAGCATTATCGTCTAGATG-3' (RNH1-Y-R)	This paper