ISCI, Volume *21*

Supplemental Information

G-quadruplex Structures Contribute to Differential Radiosensitivity of the Human Genome

Nitu Kumari, Supriya V. Vartak, Sumedha Dahal, Susmita Kumari, Sagar S. Desai, Vidya Gopalakrishnan, Bibha Choudhary, and Sathees C. Raghavan

Figure S1

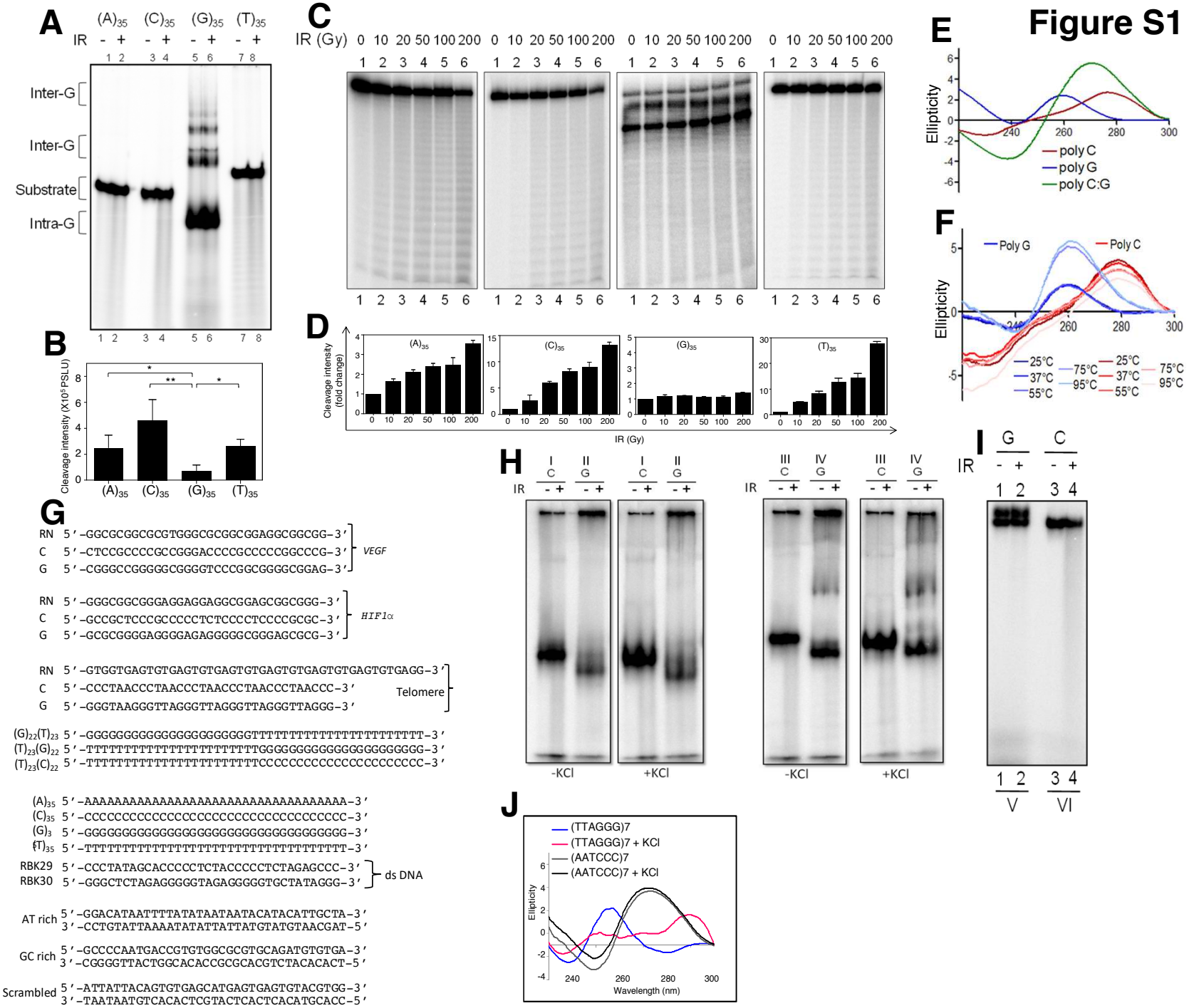
A
(A)35 (C)35 (G)35 (T)35
IR - + - + - + - +
1 2 3 4 5 6 7 8
Inter-G
Inter-G
Substrate
Intra-G
B
Cleavage intensity (X10^5 PSLU)
(A)35 (C)35 (G)35 (T)35
C
IR (Gy) 0 10 20 50 100 200
D
Cleavage intensity (fold change)
IR (Gy)
E
Ellipticity
poly C
poly G
poly C:G
F
Poly G
Poly C
Ellipticity
25°C 37°C 55°C 75°C 95°C
G
RN 5′-GGCGCGGCGCGTGGGCGCGGCGGAGGCGGCGG-3′
C 5′-CTCCGCCCCGCCGGGACCCCGCCCCCGGCCCG-3′
G 5′-CGGGCCGGGGGCGGGGTCCCGGCGGGGCGGAG-3′
VEGF
RN 5′-GGGCGGCGGGAGGAGGAGGCGGAGCGGCGGG-3′
C 5′-GCCGCTCCCGCCCCCTCTCCCCTCCCCGCGC-3′
G 5′-GCGCGGGGAGGGGAGAGGGGGCGGGAGCGCG-3′
HIF1α
RN 5′-GTGGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGG-3′
C 5′-CCCTAACCCTAACCCTAACCCTAACCCTAACCC-3′
G 5′-GGGTAAGGGTTAGGGTTAGGGTTAGGGTTAGGG-3′
Telomere
(G)22(T)23 5′-GGGGGGGGGGGGGGGGGGGGGGTTTTTTTTTTTTTTTTTTTTT-3′
(T)23(G)22 5′-TTTTTTTTTTTTTTTTTTTTTTTTGGGGGGGGGGGGGGGGGGGG-3′
(T)23(C)22 5′-TTTTTTTTTTTTTTTTTTTTTTTTCCCCCCCCCCCCCCCCCCCC-3′
(A)35 5′-AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA-3′
(C)35 5′-CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC-3′
(G)3 5′-GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG-3′
(T)35 5′-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-3′
RBK29 5′-CCCTATAGCACCCCCTCTACCCCCTCTAGAGCCC-3′
RBK30 5′-GGGCTCTAGAGGGGGTAGAGGGGGTGCTATAGGG-3′
ds DNA
AT rich 5′-GGACATAATTTTATATAATAATACATACATTGCTA-3′
3′-CCTGTATTAAAATATATTATTATGTATGTAACGAT-5′
GC rich 5′-GCCCCAATGACCGTGTGGCGCGTGCAGATGTGTGA-3′
3′-CGGGGTTACTGGCACACCGCGCACGTCTACACACT-5′
Scrambled 5′-ATTATTACAGTGTGAGCATGAGTGAGTGTACGTGG-3′
3′-TAATAATGTCACACTCGTACTCACTCACATGCACC-5′
H
I C II G III C IV G
IR - + - + - + - +
-KCl +KCl -KCl +KCl
I
G C
IR - + - +
1 2 3 4
1 2 V 3 4 VI
J
(TTAGGG)7
(TTAGGG)7 + KCl
(AATCCC)7
(AATCCC)7 + KCl
Ellipticity
Wavelength (nm)

Figure S1. Comparison of IR sensitivity in homopolymers of A, C, G and T and various DNA sequences that can fold into G4 DNA structures (Related to Figure 1). **A.** Native PAGE profile showing DNA fragmentation on homopolymeric A, C, G and T sequence, following irradiation (100 Gy). DNA substrates were irradiated and resolved on a 15% native gel. In each case, control substrates are denoted by -IR, while irradiated samples are denoted as +IR. Brackets indicate substrate in case of all four nucleotides. Slow and fast migrating species showing intermolecular (inter-G) and intramolecular (intra-G) quadruplex structures, respectively, have also been indicated. **B.** Quantification of cleavage intensity (observed as a smear), obtained from three independent repeats and plotted as a bar graph showing mean ± SEM (ns: not significant, *p < 0.05, **p < 0.005, ***p < 0.0001). **C.** Radiation-induced DNA strand-breaks when homopolymeric DNA (35 nt) of A, C, G and T sequence was exposed to increasing dose of IR. Oligomeric DNA was irradiated (0, 10, 20, 50, 100, 200 Gy), and resolved on a denaturing PAGE (15%). **D.** Each experiment was repeated a minimum of three times, the cleavage intensity was quantified using MultiGauge V3.0 and presented as a bar graph in the lower panel, showing mean ± SEM. **E.** Circular dichroism spectra of homopolymeric G, homopolymeric C and duplex C:G DNA acquired at 37^{o}C. The spectrum is plotted as a function of wavelength on the X-axis, and ellipticity values on the Y-axis **F.** Circular dichroism spectra of homopolymeric G and C sequences recorded at various temperatures (25, 37, 55, 75 and 95^{o}C). 35 mer poly Guanine (shades of blue) and poly Cytosine (shades of red) profiles are shown. The spectrum is plotted as a function of wavelength on the X-axis, and ellipticity values on the Y-axis. **G.** Oligomeric sequences of different DNA substrates used for the study. For each gene substrate, 'G' represents the G-quadruplex forming oligomer, 'C' represents its complementary region and 'RN' represents the random sequence of that particular region. Sequences used for homopolymeric A, C, G and T along with heteropolymeric (G22T23, T23G22 and T23C22) studies are also shown. Sequence of AT rich, GC rich and scrambled double-stranded DNA is also shown. **H-I.** Gel shift assay showing profiles of oligomers harboring G4 motifs from *VEGF* (I and II), *HIF1α* (III and IV) promoters (H) and telomeric region (I), upon irradiation. DNA substrates were exposed to 100 Gy IR, followed by resolution on Native PAGE (15%), in absence and presence of KCl in the gel. Corresponding complementary strands were used as controls and resolved in a similar manner. **J**. Circular dichroism spectra of oligomeric DNA harbouring human telomeric sequence. Spectrum for $(TTAGGG)_7$ in absence (blue) and presence of KCl (pink) has been shown, along with complementary control oligomers in grey (-KCl) and black (+KCl).

Figure S2

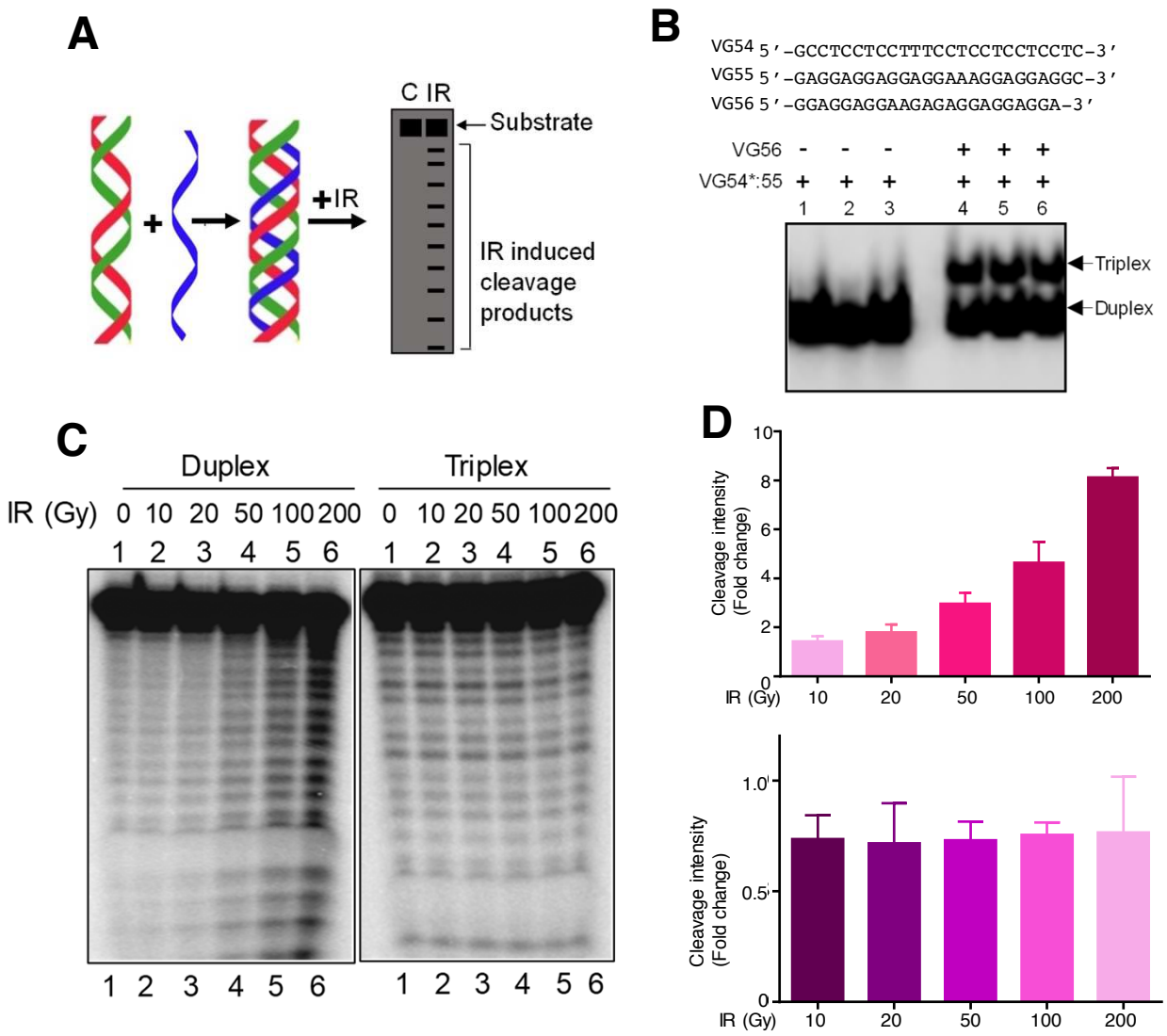
A
C IR
Substrate
+IR
IR induced cleavage products
B
VG54 5′-GCCTCCTCCTTTCCTCCTCCTCCTC-3′
VG55 5′-GAGGAGGAGGAGGAAAGGAGGAGGC-3′
VG56 5′-GGAGGAGGAAGAGAGGAGGAGGA-3′
VG56 - - - + + +
VG54*:55 + + + + + +
1 2 3 4 5 6
Triplex
Duplex
C
Duplex
Triplex
IR (Gy) 0 10 20 50 100 200 0 10 20 50 100 200
1 2 3 4 5 6 1 2 3 4 5 6
1 2 3 4 5 6 1 2 3 4 5 6
D
Cleavage intensity (Fold change)
IR (Gy) 10 20 50 100 200
Cleavage intensity (Fold change)
IR (Gy) 10 20 50 100 200

Figure S2. Comparison of IR sensitivity in oligomeric DNA that can fold into triplex DNA (Related to Figure 2)**. A.** Schematic representation of the assay employed for evaluation of IR-induced DNA strand breaks on triplex DNA structures. Radiolabeled VG54 was annealed with VG55 to form duplex DNA. Third strand, VG56 was then added and incubated at 37°C for 2 h in appropriate buffer to generate triplex DNA (A). **B.** Native PAGE profile showing formation of triplex DNA. Sequence of triplex forming oligomer used for the study is also shown. Lanes 1-3 indicates duplex DNA and lanes 4-6 is triplex DNA. **C.** Denaturing PAGE showing different forms of DNA, exposed to IR (10, 20, 50, 100 and 200 Gy), for analyzing the abundance of DNA breaks. **D.** Bar diagram showing quantification of IR induced cleavage on duplex and triplex DNA. In each case, the intensity of IR induced cleavage was subtracted from respective unirradiated control and plotted. Quantification of cleavage intensity, obtained from three independent repeats and plotted as a bar graph showing mean ± SEM.

Figure S3

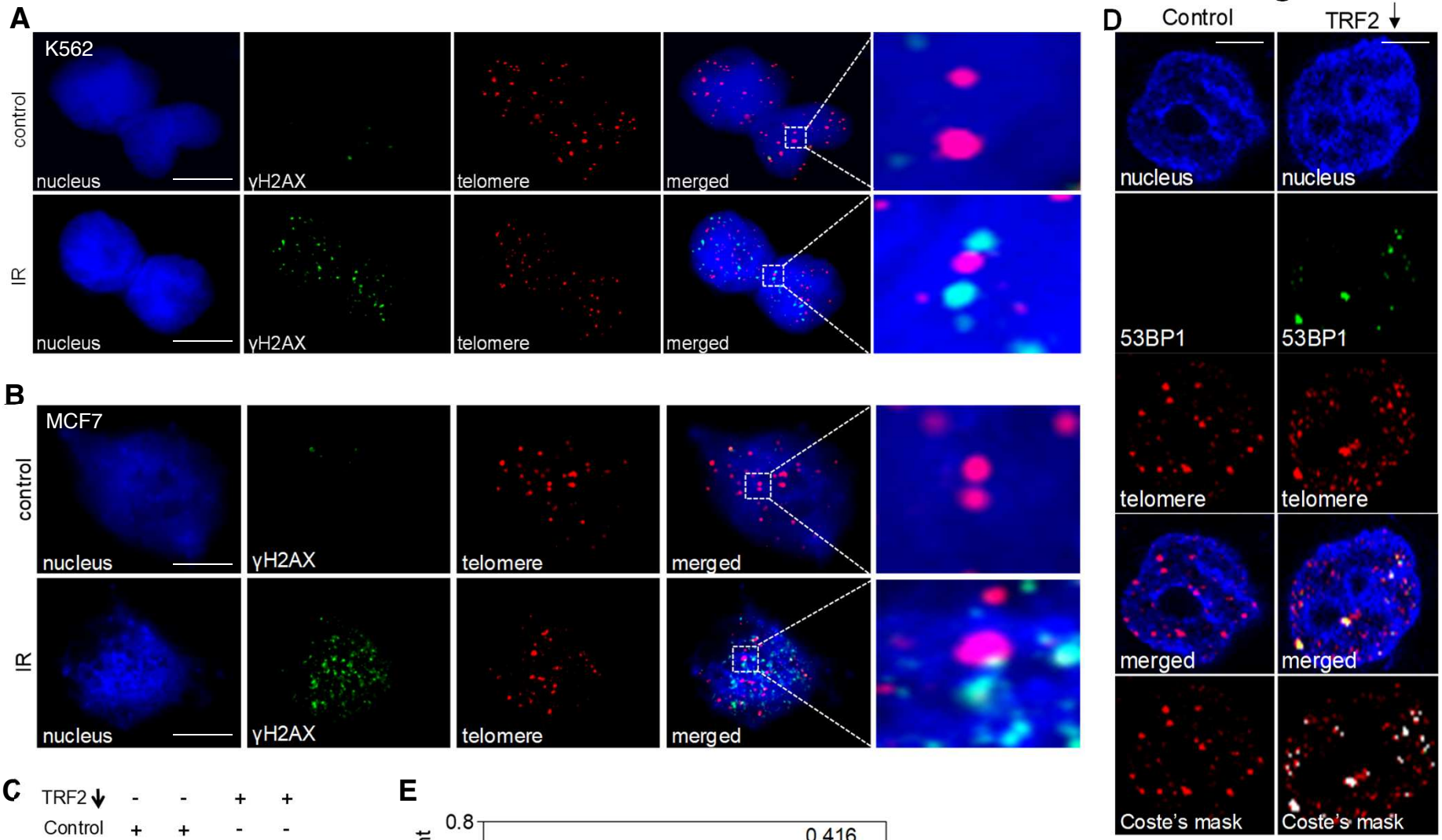
A
K562
control
IR
nucleus
γH2AX
telomere
merged
B
MCF7
C
TRF2
Control
IR
1
2
3
4
PCNA
Ponceau stained
D
Control
TRF2
53BP1
Coste's mask
E
Mander's coefficient
0.8
0.6
0.4
0.2
0.0
0.016
0.012
0.268
0.416
SCR
TRF2

Figure S3. Immunofluorescence and FISH analysis to investigate IR induced DSBs at telomere region in K562 and MCF7 cells (Related to Figure 3)**. A.** K562 cells were irradiated (10 Gy) and immunostained for γH2AX (FITC; green), followed by telomere FISH (Cy3; red). Nucleus was stained using DAPI (blue) and the merged image is shown as the extreme right panel (scale bar, 5 μm). **B.** IF- FISH analysis to investigate IR (10 Gy) induced DSBs at telomere region in MCF7 cells (scale bar, 5 μm). **C.** Representative western blot showing level of TRF2 in HeLa cells following its knockdown. A plasmid with scrambled sequence was used as control. Ponceau stained blot and level of PCNA served as loading control (scale bar, 5 μm). **D.** Representative IF-FISH images following knockdown for TRF2 in HeLa cells. Cells were transfected with either scrambled plasmid or TRF2 shRNA plasmid, and used for IF to detect 53BP1 foci (green; Alexa fluor 488). FISH was used to detect telomeres (red) and DAPI staining for nucleus (blue) (scale bar, 5 μm). **E.** Box-and-whiskers plots depicting Mander's colocalization coefficient (range: 0-1) as evaluated by JACoP plugin of ImageJ software. Experiment was performed in HeLa cells as described above, quantified and presented.

Figure S4

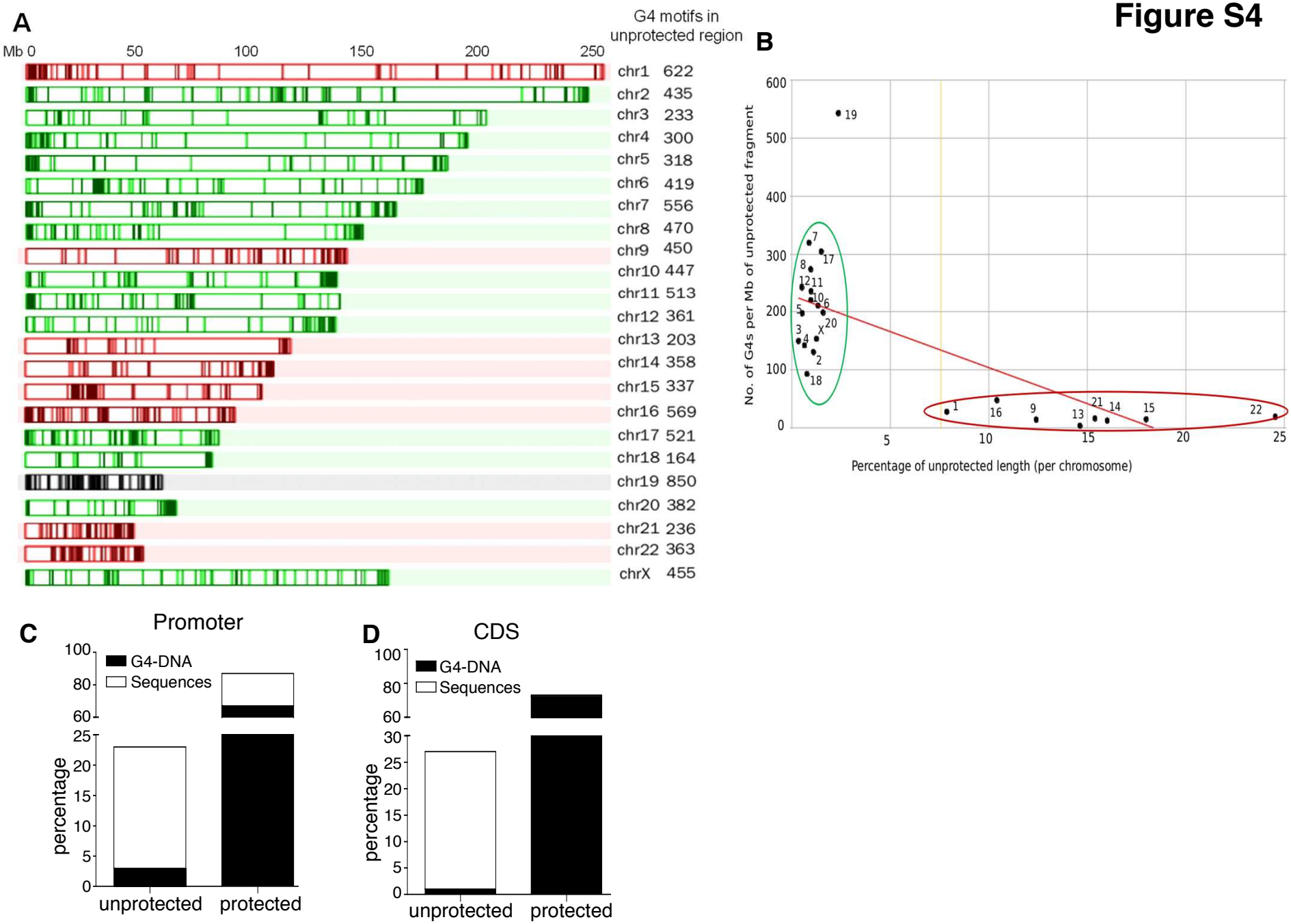
A
G4 motifs in unprotected region
Mb 0
50
100
150
200
250
chr1 622
chr2 435
chr3 233
chr4 300
chr5 318
chr6 419
chr7 556
chr8 470
chr9 450
chr10 447
chr11 513
chr12 361
chr13 203
chr14 358
chr15 337
chr16 569
chr17 521
chr18 164
chr19 850
chr20 382
chr21 236
chr22 363
chrX 455
B
No. of G4s per Mb of unprotected fragment
Percentage of unprotected length (per chromosome)
C
Promoter
G4-DNA
Sequences
percentage
unprotected
protected
D
CDS
G4-DNA
Sequences
percentage
unprotected
protected

Figure S4. Assessment of genome sequence in a human cell line for its sensitivity to IR in the context of G-quadruplex structures (Related to Figure 4)**. A.** Karyogram showing 1% damaged regions, wherein ones that are relatively less damaged are colored green and the ones that are relatively more damaged are colored red. The ones in black are outliers. **B.** A scatter plot depicting the inverse correlation between percentages of unprotected length in each chromosome vs. the number of G4 motifs per unprotected region. Higher number of G4 motifs correlates with less percentage of breakage and vice versa. The red cluster shows the chromosomes with higher number of G4 motifs and relatively less damage and the green cluster shows lower number of G4 motifs with relatively more damage. **C, D.** Histograms showing analysis of randomly selected genes for radiosensitivity (C). This is a representative set of 100 genes in which the promoter and the coding regions were assessed for G4 motifs. The graph in the right panel shows number of G4 motifs harbored by genes under study (D), both in the promoter and coding regions, which was high in the former case.

Figure S5

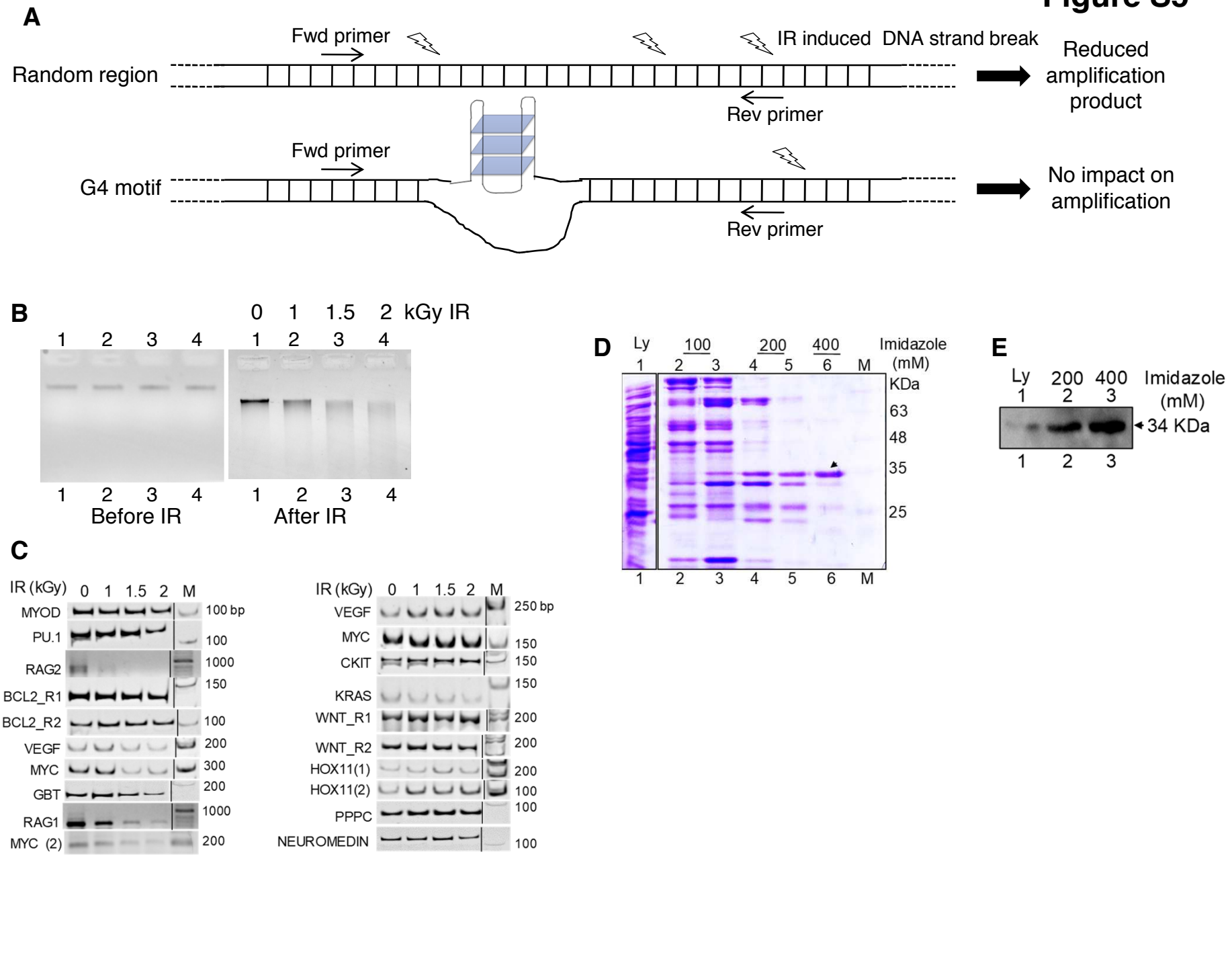
A
Fwd primer
IR induced DNA strand break
Random region
Reduced amplification product
Rev primer
Fwd primer
G4 motif
No impact on amplification
Rev primer
B
0 1 1.5 2 kGy IR
1 2 3 4
1 2 3 4
1 2 3 4
1 2 3 4
Before IR
After IR
C
IR (kGy) 0 1 1.5 2 M
MYOD
PU.1
RAG2
BCL2_R1
BCL2_R2
VEGF
MYC
GBT
RAG1
MYC (2)
100 bp
100
1000
150
100
200
300
200
1000
200
IR (kGy) 0 1 1.5 2 M
VEGF
MYC
CKIT
KRAS
WNT_R1
WNT_R2
HOX11(1)
HOX11(2)
PPPC
NEUROMEDIN
250 bp
150
150
150
200
200
200
100
100
100
D
Ly
100
200
400
Imidazole (mM)
1 2 3 4 5 6 M
KDa
63
48
35
25
1 2 3 4 5 6 M
E
Ly
200
400
Imidazole (mM)
1 2 3
34 KDa
1 2 3

Figure S5. Evaluation of randomly selected genes from human genome for their sensitivity to radiation (Related to Figures 5, 6)**. A.** Experimental strategy for detection of DNA strand-breaks upon irradiation in random regions of genome. Genomic DNA from Nalm6 cells was equalized and exposed to increasing dose of IR (1, 1.5 and 2 kGy), followed by amplification of various regions using specific forward and reverse primers. Twelve regions known to form G4 structures were amplified, along with appropriate random control regions and the DNA strand-break intensity was assessed using genomic PCR. Induction of DNA breaks in the template strand would lead to decreased amplification owing to the reduced template in the reaction, whereas an intact region would show no difference between the control and irradiated samples. **B.** Agarose gels showing equal genomic DNA in four samples before irradiation (left panel). The samples were exposed to increasing dose of IR (0, 1, 1.5 and 2 kGy) and resolved on an agarose gel (right panel). **C.** Gel profile showing genomic PCR products for control and irradiated samples (1, 1.5 and 2 kGy) for G4 forming regions, *VEGF, MYC, CKIT, KRAS, WNT*(1)*, WNT*(2)*, HOX11*(1)*, HOX11*(2)*, PPPC* and *NEUROMEDIN* (right panel), and regions devoid of G4 motifs, *MYOD, PU.1, RAG2, BCL2*(1)*, BCL2*(2)*, VEGF, MYC, GBT, RAG1* and *MYC*(2) (left panel). Molecular size marker is also shown. Vertical line indicates that image was cropped. Refer also Figure 5B. **D.** CBB gel profile showing purity of BG4 protein. Different concentrations (100-400 mM) of imidazole was used for the elution of BG4 protein that was bound to Ni-NTA column. Arrow indicates the band corresponding to the purified BG4. **E**. Western blotting to confirm the presence of BG4 protein by using anti-FLAG antibody. 'Ly' denotes total lysate.

Figure S6

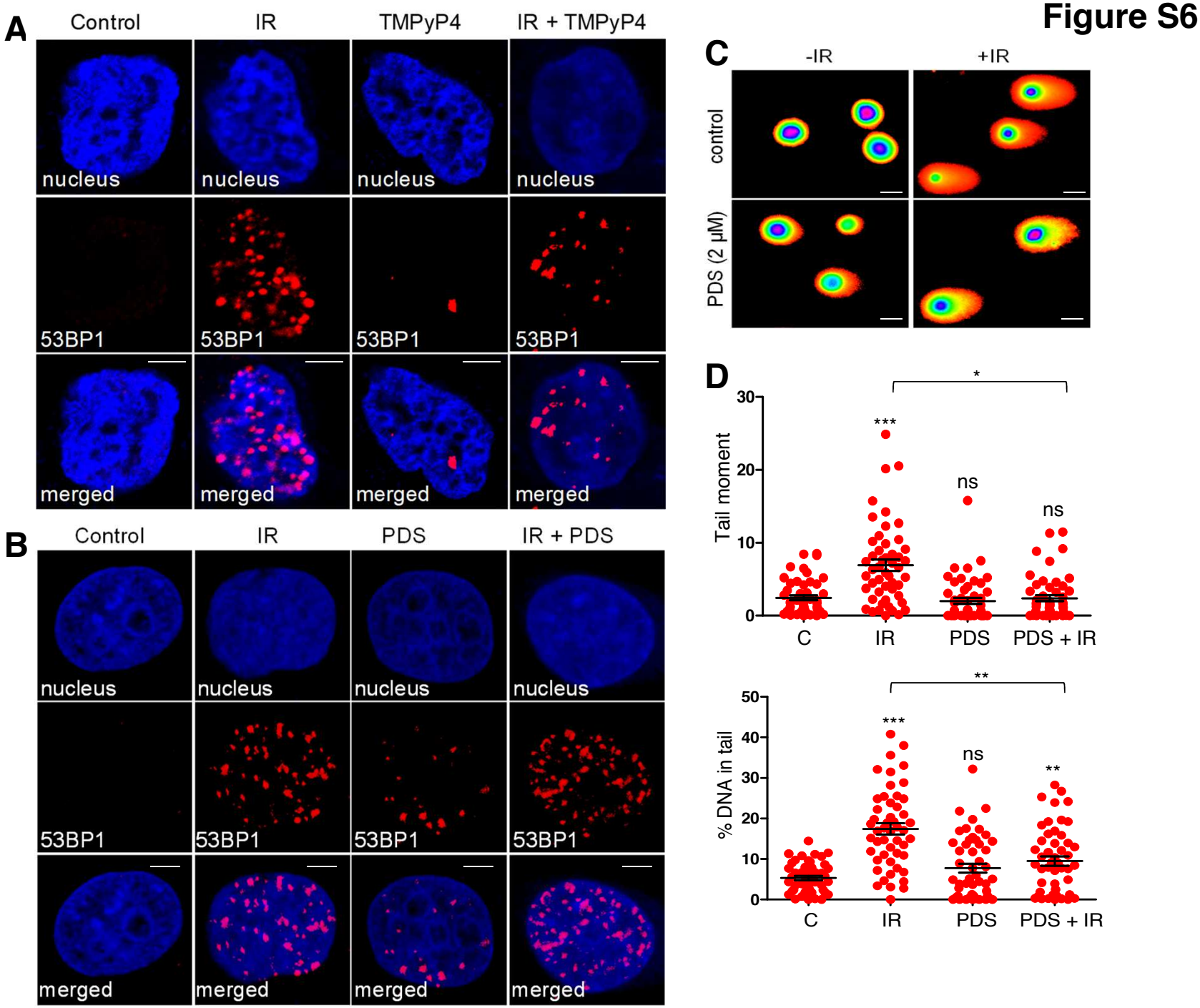
A
Control
IR
TMPyP4
IR + TMPyP4
nucleus
53BP1
merged
B
Control
IR
PDS
IR + PDS
nucleus
53BP1
merged
C
-IR
+IR
control
PDS (2 μM)
D
Tail moment
30
20
10
0
***
ns
ns
*
C
IR
PDS
PDS + IR
% DNA in tail
50
40
30
20
10
0
***
ns
**
**

Figure S6. Assessment of IR induced breaks following stabilization of G-quadruplex within the cells (Related to Figure 6)**. A.** Evaluation of DSBs following irradiation (10 Gy; 30 min recovery period) of MCF7 cells treated with a G4 stabilizing agent, TMPyP4 (5 μM, 5 h prior to irradiation). Cells alone, IR alone and TMPyP4 alone served as controls. In each case, 53BP1 foci (red, Alexa fluor 594), nucleus (blue, DAPI) and the merged image consisting of both the channels are shown (scale bar, 5 μm). **B.** Representative images showing the effect of pyridostatin (PDS) on radiation-induced DNA strand-break in HeLa cells (30 min post irradiation). Cells were treated with PDS (2 μM; 5 h), irradiated (10 Gy; 30 min recovery period) and processed for immunofluorescence as above (scale bar, 5 μm). **C, D.** Assessment of DNA breaks using comet assay, following treatment with G4 stabilizer PDS. Representative comet assay images showing control and irradiated (12 Gy; immediate harvesting) samples in presence of, pyridostatin (2 μM for 24 h), (scale bar, 10 μm) (C). Images were analyzed by using an automated software CometScore, and various parameters were calculated for each sample. Scatter plot showing quantitative parameters such as percentage DNA in tail and tail moment in each cell, and presented as mean ± SEM (D) (ns: not significant, *$p < 0.05$, **$p < 0.005$, ***$p < 0.0001$).

Figure S7

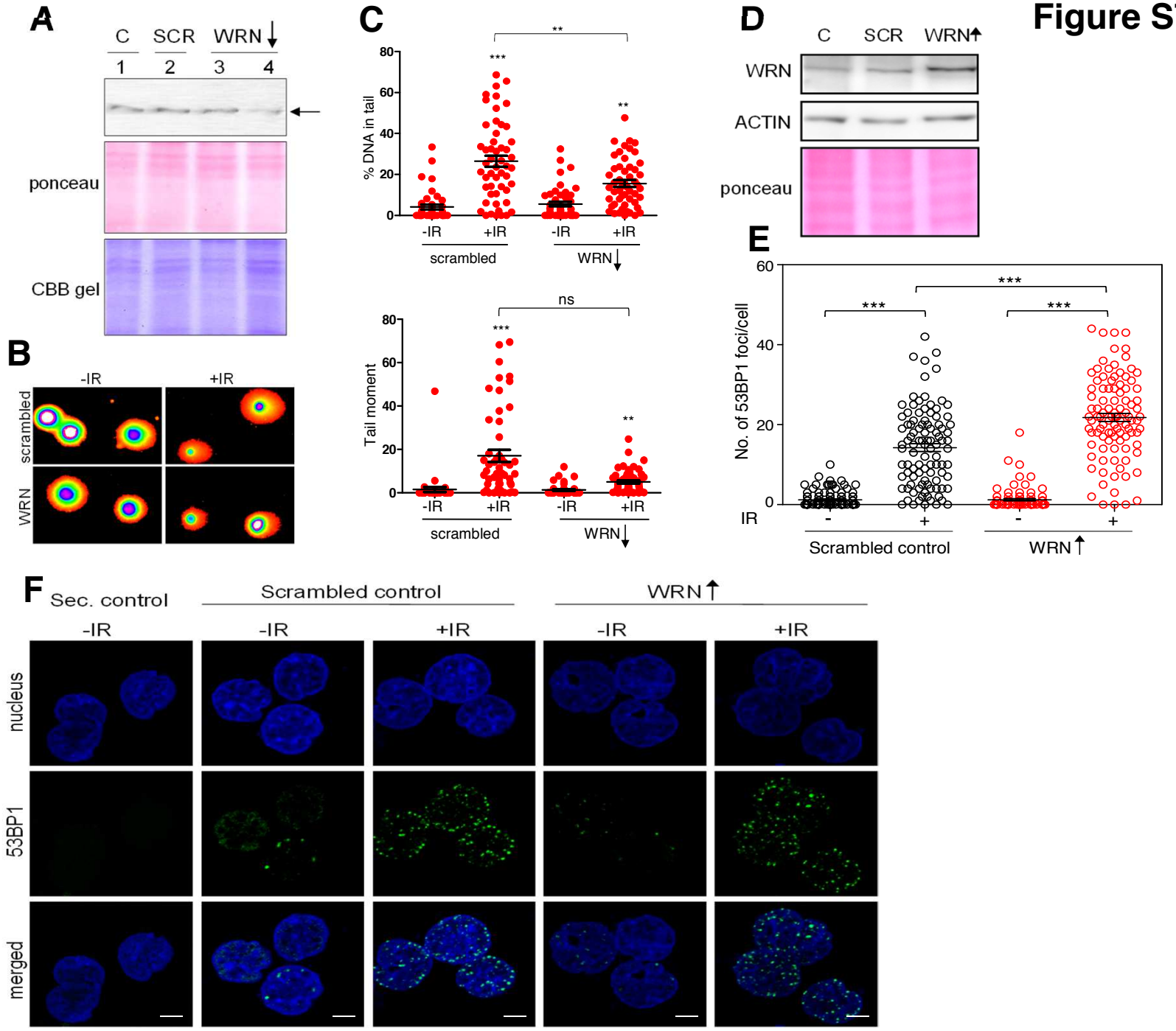
A
C
SCR
WRN
1
2
3
4
ponceau
CBB gel
B
-IR
+IR
scrambled
WRN
C
% DNA in tail
Tail moment
-IR
+IR
scrambled
WRN
ns
D
C
SCR
WRN
WRN
ACTIN
ponceau
E
No. of 53BP1 foci/cell
IR
Scrambled control
WRN
F
Sec. control
Scrambled control
WRN
-IR
+IR
nucleus
53BP1
merged

Figure S7. Evaluation of IR induced breaks following modulation of G-quadruplex formation within the cells by downregulation or upregulation of level of WRN (Related to Figure 7)**. A.** Representative western blot showing level of WRN helicase in Nalm6 cells upon transient transfection with WRN shRNA plasmids. A plasmid with scrambled sequence was used as control. Ponceau stained blot, and CBB stained gel reveal equal loading of samples. **B.** Images showing comet assay performed in cells following irradiation (10 Gy), after WRN knockdown (for 48 h) (scale bar, 10 μm). **C.** Scatter plot showing percentage DNA tail and tail moment for control and treated in knockdown samples. Comet images were analyzed using CometScore software as described above, and a scatter plot depicting mean ± SEM is shown (ns: not significant, *$p < 0.05$, **$p < 0.005$, ***$p < 0.0001$). **D.** Representative western blot showing overexpression of WRN helicase in Nalm6 cells. Actin and ponceau stained blot served as loading controls. 'C' refers to cells alone control, whereas 'SCR' denotes the scrambled plasmid control for transfection. **E**. Scatter plot showing quantification of number of 53BP1 foci upon irradiation following WRN overexpression, as compared to transfection control. A minimum of 100 cells were analyzed for each sample, 53BP1 foci counted and plotted as scatter plot using GraphPad Prism 5 software depicting mean ± SEM (ns: not significant, *$p < 0.05$, **$p < 0.005$, ***$p < 0.0001$). **F**. Representative immunofluorescence images showing 53BP1 foci formation (green) upon irradiation following WRN overexpression. Nucleus was stained with DAPI (blue), and merged image of both is shown in the lower column. Post transfection, cells were irradiated (10 Gy), allowed to recover for a period of 30 min (scale bar, 2 μm).

Figure S8

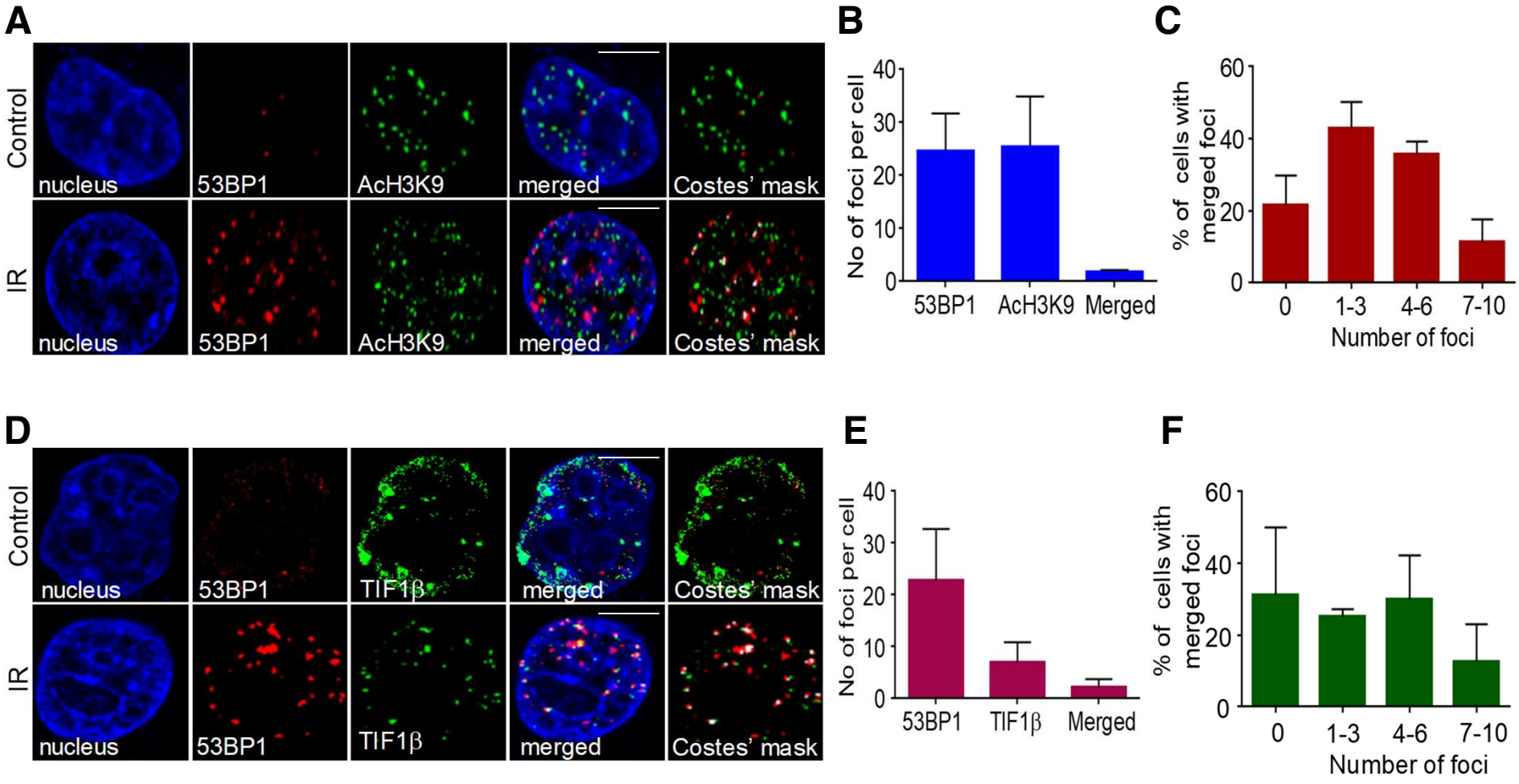
A
Control
IR
nucleus
53BP1
AcH3K9
merged
Costes' mask
B
No of foci per cell
40
30
20
10
0
53BP1
AcH3K9
Merged
C
% of cells with merged foci
60
40
20
0
0
1-3
4-6
7-10
Number of foci
D
Control
IR
nucleus
53BP1
TIF1β
merged
Costes' mask
E
No of foci per cell
40
30
20
10
0
53BP1
TIF1β
Merged
F
% of cells with merged foci
60
40
20
0
0
1-3
4-6
7-10
Number of foci

Figure S8. Immunofluorescence analysis to evaluate the role of chromatin organization on induction of DNA breaks by IR (Related to Figure 7)**. A.** Representative immunofluorescence images showing control and irradiated cells (10 Gy). Cells were stained for 53BP1 and euchromatin marker, acetylated histone H3K9 (scale bar, 5 μm). **B**. Bar graph showing average number of 53BP1, AcH3K9 and the merged foci per cell upon irradiation. **C.** Bar graph showing the distribution of colocalized foci of 53BP1 and AcH3K9 in cells following irradiation. **D.** Representative immunofluorescence images showing 53BP1 and heterochromatin marker (TIF1β) stained control and irradiated cells (10 Gy) (scale bar, 5 μm). **E.** Bar graph showing average number of 53BP1, TIF1β, and the merged foci per cell. **F.** Bar graph showing the distribution of colocalized foci of 53BP1 and TIF1β in individual cells. In panels B, C, E and F quantification is based on a minimum of two hundred cells in each case and plotted as a bar graph showing mean ± SEM.

TRANSPARENT METHODS

Oligomeric DNA

Oligomeric DNA used in the study were purchased from Sigma-Aldrich (St Louis, MO, USA) or IDT (San Jose, USA). The sequence of oligomers used were MS5, 5'-AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA-3'; MS4, 5'-CCCCCCCCCCCCCCCC-CCCCCCCCCCCCCCCCCCCC-3'; MS9, 5'- GGGGGGGGGGGGGGGGGGGGGGGGGGGG-GGGGGGGG-3'; MS3, 5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-3'; MN150, 5'-GCGCGAGGGAGGGGAGGGGAGGGGGAGAGGGAAACAGGCCACGTAAAGCAACTCTCTA AAGGCAGCGTC-3'; MN151, 5'-GCGCGAGGGAATAAATAATAGGGGGAGAGGGAAACAGG-CCACGTAAAGCAACTCTCTAAAGGCAGCGTC-3'; MN154, 5'-GCGCGACTAAATAAATAAT-AGACAGAGACTAAAACAGGCCACGTAAAGCAACTCTCTAA-AGGCAGCGTC-3'; MN117, 5'-GCGGAGCGAGCCGAGCGAGCGTGTGAGGTTGAGTGCGTTGAAACAGGCCACGTAAAGCA ACTCTCTAAAGGCAGCGTC-3'; SV10, 5'-GGGGGGGGGGGGGGGGGGGGGGGTTTTTTT-TTTTTTTTTTTTTTT-3'; SV11, 5'-TTTTTTTTTTTTTTTTTTTTTTTGGGGGGGGGGGGGGGG-GGGGGGG-3'; SV12, 5'-TTTTTTTTTTTTTTTTTTTTTTTCCCCCCCCCCCCCCCCCCCCCC-3'; SV25, 5'-GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG-3'; SV26, 5'-CCCTAACCCT-AACCCTAACCCTAACCCTAACCC-3'; SV27, 5'-GGGTTAGGGTTAGGGTTAGGGTTAGGGT-TAGGGTTAGGGTTAGGG-3'; SV28, 5'-CCCTAACCCTAACCCTAACCCTAACCCTAACCC-TAACCCTAACCC-3'; VG53, 5'-GTGGTGAGTGTGAGTGTGAGTGTGAGTGTG-AGTGTGAGTGTGAGG-3'; VG17, 5'-TATAAAATGATTCTATCTAGTCTGGGGGC-3'; KD49, 5'-CGGGCCGGGGGCGGGGTCCCGGCGGGGCGGAG-3'; KD50, 5'-CTCCGCCCCGCC-GGGACCCCGCCCCCGGCCCG-3'; VG49, 5'-GGCGCGGCGCGTGGGCGCGGCGGAGG-CGGCGG-3'; MN89, 5'-GCCGCTCCCG-CCCCCTCTCCCCTCCCCGCGC-3'; RT17, 5'-GCGCGGGGAGGGGAGAGGGGGCGG-GAGCGCG-3'; VG51, 5'-GGGCGGCGGGAGGA GGAGGCGGAGCGGCGGG-3';VG54, 5'-GCCTCCTCCTTTCCTCCTCCTCCTC-3'; VG55, 5'-GAGGAGGAGGAGGAAAGGAGGAGGC-3'; VG56, 5'-GGAGGAGGAAGAGAGGAGGAGGA-3'; AKN86 (Hp1) 5'-CGGGATCGGAGGGATACGAGCTCGTATCCCTTTGATCCCG-3'; AKN87 (Hp2), 5'-CGAGATCGGAGAGATAGAAAAAACTATCTCTCCGATCTCG-3'; AKN88 (Hp3), 5'-CGAGATCGGAGAGATAGCCCCCCCTATCTCTCCGATCTCG-3'; AKN89 (Hp4), 5'-CGAGA-TCGGAGAGATAGTTTTTTCTATCTCTCCGATCTCG-3'; AKN90 (Hp5), 5'-CGAGATCGGAG-AGATAGGGGGGGCTATCTCTCCGATCTCG-3'; VG60, 5'-GGGCTGGGTGGGTGC-TGTCAAGGGCTGGG-3'; BTM2, 5'-GAAGGAGGGCAGGAGGGCTCTGGGTGGGTCTGT-3'; MN38, 5'-GCGCGAGGGAGGGGAGGGGAGGGGGAGAGG-3'; MN45, 5'-AGAAGGGG-

GAGGGGAGGGAGAGAGGGGGCGCCG-3'; DR53, 5'-GTTTTGCCCGGGGTC-CCGGGCGCTTTGGGC-3'. SK9, 5'-GGACATAATTTTATATAATAATACATACATTGCTA-3'; SK10, 5'-TAGCAATGTATGTATTATTATATAAAATTATGTCC-3'; SK11, 5'-GCCCCAATGA-CCGTGTGGCGCGTGCAGATGTGTGA-3'; SK12, 5'-TCACACATCTGCACGCGCCA-CACGGTCATTGGGGC-3'; SK13, 5'-ATTATTACAGTGTGAGCATGAGTGAGTGTACGTGG-3'; SK14, 5'-CCACGTACACTCACTCATGCTCACACTGTAATAAT-3'; SV23, 5'-GGGTTAGGGTTAGGGTTAGGG-3'; SV24, 5'- CCCTAACCCTAACCCTAACCC-3'. Telomere probe sequence, 5'-Cy3/CCCTAACCC-TAACCCTAA-3'; centromere probe sequence, 5'-Cy3/ ATTCGTTGGAAACGGGA-3'. Telomere and centromere probes were from IDT, USA.

All oligomers were gel purified on 8-15% denaturing polyacrylamide gel. The purified substrates were then radiolabeled at 5' end using γ-^{32}P-ATP, and T4 polynucleotide kinase as described before (Nambiar and Raghavan, 2012).

Cell lines and culture

MCF7 (human breast cancer), HeLa (human cervical cancer), HEK293T (human embryonic kidney cells) and K562 (human chronic myelogenous leukemia) were purchased from National Centre for Cell Science, Pune, India. Nalm6 cells were from Dr. M. R. Lieber. Cells were cultured in RPMI 1640 or MEM medium supplemented with 15% fetal bovine serum (FBS), 100 µg/ml Penicillin, and 100 µg/ml streptomycin and incubated at 37^{o}C in a humidified atmosphere containing 5% CO_2 as described before (Srivastava et al., 2012).

Irradiation

Samples were irradiated at room temperature using a Cobalt-60 gamma irradiator (BI 2000, BRIT, India). The dose rate of the source at the time of usage was 0.91 Gy/min. IR dose employed in the study varied according to the sample and the experiment, and are specified in the respective assays.

Preparation of DNA substrates

Double stranded oligomeric DNA was prepared by annealing radiolabeled MS3 and MS4 with their complementary strand MS5 and MS9, respectively in presence of 100 mM NaCl and 1mM EDTA as described before (Kumar et al., 2010). AT rich, GC rich and scrambled oligomers were prepared by annealing SK9 and SK10, SK11 and SK12, SK13 and SK14, respectively. Double stranded DNA containing telomeric repeat sequences was prepared by annealing SV23 and SV24 (3 repeats, Telo A), SV25 and SV26 (5 repeats, Telo B), SV27 and SV28 (7 repeats, Telo C), respectively.

For preparation of triplex DNA (Rustighi et al., 2002), previously described oligomers were synthesized and used. Radiolabeled oligomer, VG54 (60 nM) was annealed with unlabelled VG55 (300 nM) in presence of 100 mM NaCl and 1 mM EDTA. Third strand oligomer, VG56 was then annealed to the duplex DNA (Wu et al., 2007) in 1x triplex buffer (10 mM Tris [pH 7.5], 100 mM NaCl, 10 mM $MgCl_2$ and 1% glycerol) for 2 h at 37°C.

Plasmid constructs

WRN helicase overexpression vector, pMM290 (Addgene plasmid # 46038) was a gift from Raymond Monnat (USA) (Swanson et al., 2004).

Plasmid isolation and purification

For each plasmid, after transformation, *Escherichia coli* were cultured in 500 ml of Luria broth (Hi Media, USA) for 18 h at 37°C. Isolation of plasmid DNA was performed by standard alkaline lysis method (denaturing method) and purified by cesium chloride-ethidium bromide density gradient centrifugation as described (Sambrook et al., 1989). Briefly, 5.1 g CsCl was dissolved in 4.0 ml TE. About 5 mg of isolated plasmid and 0.7 mg Ethidium Bromide were added to CsCl solution, vortexed and centrifuged using rotor TLA110 and Quick seal tubes (Beckman Coulter, USA) for 12 h, 72,000 rpm at 20°C. Following removal of ethidium bromide by butanol extraction, the plasmid DNA was precipitated with isopropanol and was washed in 70% alcohol. The pellet was dissolved in TE (pH 8.0).

IR-induced DNA breaks

For studying the effect of IR-induced DNA breaks, radiolabeled oligomers were irradiated (10, 20, 50, 100, 200 Gy) and resolved on either denaturing or native polyacrylamide gels. In case of triplex DNA, increasing doses of radiation (10, 20, 50, 100 and 200 Gy) was used. A single dose (100 Gy) was used for irradiation of hairpin DNA substrates (Hp1-Hp5). To study the effect of irradiation induced single stranded versus double-stranded breaks, ssDNA and dsDNA substrates were irradiated with a dose of 100 Gy. To study the effect of radiation on G-rich oligomers derived from different genomic regions, *VEGF*, *Hif1α* and telomere, KD49, RT17 and SV27, respectively were radiolabeled, incubated in TE buffer in presence or absence of KCl (Nambiar et al., 2011) and irradiated (150 Gy). Complementary C-rich oligomers (KD50, MN89 and SV28, respectively) and random oligomers (VG49, VG51 and VG53) derived from same genomic regions were also used for irradiation. In all cases, following irradiation, products were resolved either on 12-15% denaturing or native polyacrylamide gels. The gels were then dried, exposed to a screen, and signals were detected using a PhosphorImager (FLA9000, Fuji, Japan). In all experiments, unirradiated oligomeric DNA served as control. Quantification of the

IR-induced DNA cleavage was carried out using Multi Gauge V3.0, wherein the unirradiated control lane was subtracted from the irradiated ones.

Genomic DNA from Nalm6 cells was exposed to increasing dose of radiation (1, 1.5 and 2 kGy), followed by PCR amplification of the gene of interest using appropriate primers.

In case of IF-FISH experiments, cells were irradiated with a dose of 10 Gy, and incubated at 37^{o}C for a 30 min or 10 min recovery period, followed by harvesting and processing for immunofluorescence/FISH/comet assay. Radiation doses used for other experiments are indicated at appropriate figure legends.

Purification of BG4 protein

The plasmid expressing BG4 protein, pSANG10-3F-BG4 was a gift from Shankar Balasubramanian (Addgene plasmid # 55756). The plasmid was transformed into *E. coli*, BL21(DE3), and the culture was expanded by incubating at 30^{o}C, till the O.D. reached upto 0.6 (Biffi et al., 2013). The protein expression was then induced with 1 mM IPTG for a period of 16 h at 16^{o}C, harvested, and resuspended in lysis buffer (20 mM Tris-HCl [pH 8.0], 50 mM NaCl, 5% glycerol, 1% Triton X-100 and 1 mM PMSF). The cells were lysed by sonication, centrifuged, and the supernatant was then loaded onto a Ni-NTA chromatography column (Novagen). BG4 was eluted using increasing concentrations of Imidazole (100-400 mM), and purity was checked by CBB staining. BG4 enriched fractions were dialyzed against dialysis buffer (PBS containing 0.05% Triton X-100, 1 mM beta-mercaptoethanol, 5% glycerol and 0.1 mM PMSF) overnight at 4^{o}C. Identity of the protein was confirmed by immunoblotting, using an anti-FLAG antibody.

Circular dichroism (CD)

In order to assess G-quadruplex formation, DNA oligomers were subjected to CD spectroscopy at room temperature using a JASCO J-810 spectropolarimeter (50 nm/min scan speed, 10 cycle accumulations, 220-300 nm scan range) as described before (Nambiar et al., 2013; Raghavan et al., 2005). DNA of interest [ss DNA, ds DNA were resuspended in TE buffer, either in the presence or absence of 100 mM KCl or LiCl at 37^{o}C for 1 h, followed by CD spectroscopy measurements. The spectrum for buffer alone was subtracted from the experimental spectra. The ellipticity was obtained using the Spectra Manager software and was plotted as a function of wavelength. For annealing of double stranded DNA, equimolar concentrations of MS4 and MS9 were boiled in presence of 100 mM NaCl and 1 mM EDTA for 10 min and allowed to anneal by slow cooling.

Gel Mobility Shift Assay

Radiolabeled oligomers were incubated in the presence of 100 mM KCl for 1 h at 37°C, followed by irradiation as described before (Nambiar et al., 2011; Nambiar et al., 2013). The oligomers were then resolved on 12-15% native polyacrylamide gels, run at 150 V at room temperature, in the presence of KCl in the gel as well as the running buffer. The gels were dried, exposed to a PhosphorImager screen, and the signals were detected using a PhosporImager FLA9000 (Fuji, Japan).

Triplex DNA was prepared as described above by annealing radiolabeled VG54 with VG55 in presence of 100 mM NaCl and 1 mM EDTA. Annealed dsDNA was then incubated with triplex forming oligomer, VG56 in appropriate buffer (1 mM Tris [pH 7.5], 10 mM NaCl, 1 mM $MgCl_2$ and 1 % glycerol) for 2 h at 37°C. The oligomers were then resolved on 12% native polyacrylamide gels, (100 V at room temperature), in the presence of 100 mM $MgCl_2$ in 1x Tris Borate (pH 7.8) running buffer and radioactive signals were detected as described above.

DMS protection assay

Radiolabeled oligomers were incubated with 100 mM KCl (wherever indicated) at 37°C for 1 h, followed by irradiation (100 Gy). 1 µl DMS (1:10 stock) was added to the 10 µl reaction mixture and incubated at room temperature for 15 min, as described before (Das et al., 2016; Nambiar et al., 2011). Equal volume of piperidine (1:10 stock) was then added and the samples were heated for 30 min at 90°C. DNA was purified, loaded onto an 18% denaturing polyacrylamide gel, followed by drying, and subsequent visualization of the bands as described above.

The oligomers derived from BCL2 major breakpoint region (BTM2) (Nambiar et al., 2011), Region I (MN38) and II (MN45) of the HOX11 breakpoint region (Nambiar et al., 2013), BU1A (VG60) (Schiavone et al., 2014) and human telomeric region (SV27) were used for the study. Oligomeric DNA containing a random sequence, VG17 was also used for the study.

Immunocytochemistry

Cells were cultured on coverslips (50,000/ml for 24 h), irradiated (10 Gy), and allowed to recover for a period of 30 min at 37°C. Control cells (without IR) were sham treated, and processed in a similar way. After the recovery period, cells were harvested, washed with phosphate buffered saline (PBS) and fixed using 2% paraformaldehyde, PFA (10 min at room temperature), as described before (John et al., 2015; Sebastian and Raghavan, 2016). Following fixing, cells were permeabilized using PBS + 0.1% Triton X-100 (5-10 min at RT), blocked using PBST + 0.1% BSA + 10% FBS (1 h at 4°C), and incubated with appropriate primary antibodies, γ-H2AX (Cell Signaling Technology, USA), pATM, 53BP1, TRF2, TIF1β,

AcH3K9 (Santa Cruz Biotechnology, USA) at 4ºC overnight. Corresponding Alexa Fluor conjugated with secondary antibodies (Life Technologies, USA) were added and the cells were incubated at 4ºC for 2 h. Following antibody staining, the coverslip was mounted using DAPI, diazobicyclo[2.2.2]octane (DABCO) mix. Images were captured using Laser Confocal microscope (Zeiss, Germany, Olympus, FLUOVIEW FV3000, Japan) or Apotome Fluorescence Microscope (Zeiss, Germany).

For BG4 and γH2AX colocalization studies, HeLa or MCF7 cells were irradiated (10 Gy) and used for immunofluorescence as described above. Besides, a radiation dose titration (5, 10 and 20 Gy) was also performed in HeLa cells. In case of BG4, anti-FLAG was used for detection.

When required, images were analyzed for colocalization of signals by using the JACoP plugin of the ImageJ software, as described previously (Bolte and Cordelieres, 2006; Kumar et al., 2010). Mander's colocalization coefficient, which denotes the fraction of overlap of one color over the other, was chosen as one of the parameters of analysis. The value obtained for each cell was plotted in the form of a box and whisker plot, denoting mean ± SEM.

IF-FISH

Cells (MCF7, K562 and HeLa) were seeded, irradiated and processed for immunocytochemistry, as described above. Following IF, FISH was performed as described previously, with minor modifications (Dimitrova et al., 2008). Briefly, the cells were fixed post-IF using 2% PFA (2 min) followed by RNase A treatment (0.1 mg/ml; 1 h at 37ºC). The sample was washed twice with 2XSSC, once with distilled water and incubated with trypsin (5%) for 4 min at 37ºC. The sample was fixed again (2% PFA, 2 min, RT), followed by serial dehydration in 70, 90 and 100% ethanol (5 min each). Following air drying, the sample was denatured (85ºC for 5 min), incubated with the telomere or centromere probes (IDT, USA) for 10 min at 85ºC, and allowed to hybridize at room temperature (2 h). The coverslip was then mounted using DAPI:DABCO mix.

IF-FISH was also performed following knockdown of TRF2 in HeLa cells. For this, cells growing in log phase ($5x10^5$ cells) were transfected with shRNA plasmid (9 µg) that can target TRF2 (TRCN0000004811, shRNA core facility of IISc) using polyethylenimine method (Iyer et al., 2016). Cells were harvested after 24 h of incubation (37ºC) and TRF2 expression level was assessed by western blotting. For investigating the radiation induced DNA breaks, following TRF2 knockdown, cells were irradiated (10 Gy), allowed to recover for 30 minutes. Irradiated

cells were then subjected to immunofluorescence for detection of 53BP1 foci, following which IF-FISH was performed as described above.

In all cases, images were analyzed for colocalization of signals by using the JACoP plugin of the ImageJ software, as described previously (Bolte and Cordelieres, 2006; Kumar et al., 2010). Mander's colocalization coefficient, which denotes the fraction of overlap of one color over the other, was chosen as one of the parameters of analysis. The value obtained for each cell was plotted in the form of a box and whisker plot, denoting mean ± SEM.

Genomic DNA Extraction

Genomic DNA was extracted from Nalm6 cells, as described before (Nambiar and Raghavan, 2010; Raghavan et al., 2004). Briefly, cells were lysed with extraction buffer in the presence of Proteinase K (overnight at 37^{o}C) and genomic DNA was purified following phenol-chloroform extraction. DNA was resuspended in TE buffer and stored at 4^{o}C until use.

Real time PCR of irradiated genomic DNA

Genomic DNA was irradiated using gamma rays (1, 1.5 and 2 kGy) after resuspending in TE. Real time PCR was performed on irradiated genomic DNA samples in a BioRad IQ5 Real time PCR Detection System Ver2.1. Unirradiated DNA served as the control. Briefly, 10 ng of genomic DNA, 5 µl of 2x SYBR Green master mix containing dNTPs, $MgCl_2$ and Taq DNA Polymerase (Takara Bio Inc.) and 500 nM each of the respective forward and reverse primer were mixed in a total reaction volume of 10 µl. The reaction was run at 95^{o}C for 3 min, followed by 35 cycles of 95^{o}C for 30 sec, 58^{o}C for 30 sec and 72^{o}C for 30 sec. All PCRs were performed in triplicates. Ct values for respective control and irradiated samples were determined using Bio-Rad IQ5 analysis software.

Real time PCR primers used for the study were SV42, 5'-TGGGGGAGCGTGTCAGAAT-3', SV43, 5'-GCTTTAATTGTGTGATTGGAC-3' for *VEGF* G4, SV50, 5'-AGTTCCCTGGCAACATCTG-3', SV51, 5'-AATTTGGCACCAAGTTTGT-3' for VEGF random, SV46, 5'-GGAGGAGCAGCAGAGAAAG-3', SV47, 5'- GTGGGGAGGGTGGGGAAGGT-3' for *MYC* G4, SV52, 5'-TAGGCTGGAGGTCGTGGTTA-3', SV53, 5'-CGGCGCTTTCGGATTAACT-3' for *MYC* random, SV48, 5'- AGCAAGTAGTAATTGATGG-3', SV49, 5'-GCAAATACACAAAGAAAGC-3' for *KRAS* G4, SV54, 5'-ACACAGTCCAGACACTCTGC-3', SV55, 5'-ACGTGCAGAAACTCCTTGTTC-3' for *WNT* G4 region 1, SV56, 5'-CCAGCGCCGCAACTATAAGA-3', SV57, 5'-GGCGACTTTGGTTGTTGCCC- 3' for *WNT* G4 region 2, SV60, 5'-TAAGGCACTTTCTGGGCAGT-3', SV61, 5'-CTGCCAACCCCAGTAATGAT-3' for *CKIT* G4, SV66, 5'-GCAGGATAGCAGCACAGGATT-3',

SV67, 5’-GGCCTCAGGGAACAGAATGAT-3’ for BCL2 random region 2, SV58, MN34, 5' TTAAGCCTCGCCTTGTTC-3', MN35, 5'- CGGTGCAAGAGAGCTTCG-3' for HOX11, 5’-CAGTCTTCAGGCAAAACGTCGA-3’, SV59, 5’-TGGTCGGATTTCCAAAGACA-3’ for BCL2 random region 1, SV68, 5’-CCTCTTTCGGTCCCTCTTTC-3’, SV69, 5’-ATGGGTAGAGCGGCTGTAGA-3’ for MYOD random region, SV70, 5’-TGGTGGATAGGCAAGAAAGG-3’, SV71, 5’-ATCCCTAGGGCTCTGTTTCC-3’ for PU.1 random region, SV80, 5'-AAAACAACTTCCACAGTCCT-3', SV81, 5'-CAAGCTGGTTACGCCCCAGA-3' for PPPC G4, SV86, 5'-GCCCCGCGTTCTCCG-3', SV87, 5'-ACACACTCGGCCATTACGG-3' for USLP-1 G4, SV90, 5'-AACGGCCTCGGAGAAGTAAC-3', SV91, 5'-TGGCGCATTCGACCCAAATA-3' for NEUROMEDIN G4, SV92, 5'-GTCGGACTCCTACCCCTTTTG-3', SV93, 5'-CTGGTTTAACAGGCGTTCCC-3' for CALNEXIN G4, SV94, 5'-GGTCTTGTTCCTAAGGGGGC-3', SV95, 5'-TTTGACTGCGTGTTGTGCAG-3' for MYC random region 1, SV100, 5'-GGAATCCACCTCTTGAAGGCA-3', SV101, 5'-CTCCGCTCTGACCCAAGAAC-3' for GBT1 random region, SV104, 5'-AAGCTGGGGCTTATACTGACTG-3', SV105, 5'-GAGAGTTTGCTGGGGCACTG-3' for RAG1 random region, SV106, 5'-ACCCTCGTAGCTCGCACTTA-3', SV107, 5'-GGCTAGTCCGAAGGTGCAAA-3' for MYC random region 2, AP21, 5’-CTGCCCTAC-TTGTGATGTGG-3’, AP22, 5’-GATGGATGAGTGTGCGTTCT-3’ for RAG2, MN91, 5'-GAA-GCTCTCTTGCACCG-3', SKS10, 5'-CAAGGATCCGCCAACAGCGGCTCCTGGC-3' for HOX11 region II. Same conditions and primers were also used for semiquantitative PCR using γ-irradiated genomic DNA. The additional primers used for semiquantitative PCR were AP1, 5'-GCTGATGGTCAGAAGCCAGT-3', AP2, 5'-GTTGGTCTCCACAGGGTCAG-3', for RAG1, AP3, 5'-AGCCAGGCTTCTCACTGATG-3', AP4, 5'-GGTTGGCAGGCCGGATATTA-3' for RAG2.

Neutral comet assay

Generation of DNA double-strand breaks following irradiation was determined using the comet assay, as described previously (John et al., 2015; Sebastian and Raghavan, 2016). Briefly, cells were harvested post irradiation and washed using PBS. ~10,000 cells were mixed with low melting agarose and spread on an agarose coated glass slide. Cells were lysed overnight at 37^{o}C using neutral lysis buffer (0.5 M EDTA, 2% sarkosyl and 0.5 mg/ml Proteinase K). Slides were then rinsed thoroughly in neutral electrophoresis buffer (90 mM Tris HC (pH 8.0), 90 mM Boric Acid and 2 mM EDTA), and electrophoresed at 12 V for 25 min. Slides were rinsed in double-distilled water, and stained using propidium iodide (2.5 µg/ml for 20 min).

Comets were imaged using an Apotome Fluorescence Microscope (Zeiss, Germany). Comet analyses for the % DNA in tail, tail moment and olive moment were performed using the CometScore software as described before (John et al., 2015; Sebastian and Raghavan, 2016), and the values were plotted as a scatter plot showing individual cell analysis.

G-quadruplex stabilization using small molecules

HeLa and MCF7 cells (50,000/ml) were treated with TMPyP4 (5 μM) for 5 h followed by 10 Gy irradiation. After recovery periods of 30 min, cells were harvested and processed for immunofluorescence against 53BP1 protein, as described earlier. G-quadruplex stabilization by Pyridostatin was performed by treating cells with the compound (2, 5 μM, 24 h), followed by irradiation (10 Gy). Cells were harvested after 30 min and IF was performed using anti-53BP1. In experiments where neutral comet assay was performed to detect DSBs, cells were harvested immediately and used as described above.

WRN helicase knockdown and overexpression within cells

WRN helicase knockdown (using shRNA plasmid against WRN, TRCN-AAB87, B and C), and overexpression (overexpression vector, pMM290) was performed by using the polyethylenimine method of transfection, as described before (Iyer et al., 2016). In brief, ~ $5x10^5$ lakh cells (Nalm6, and HeLa) were transfected with 10 μg of the plasmid, followed by 24 h (for knockdown) or 36 h (for overexpression) incubation at 37^{o}C. Cells were harvested and WRN protein levels were assessed using western blotting. For assessing the effect of radiation on these samples, cells were irradiated (10 Gy), allowed to recover for 30 minutes (or harvested immediately in case of comet assay) and was subjected to either immunofluorescence for detection of 53BP1 foci or neutral comet assay, as described above. In case of the samples where WRN knockdown was done followed by overexpression, cells were first transfected with shRNA (37^{o}C for 24 h) followed by transfection with overexpression construct, pMM290 (37^{o}C for 36 h) and WRN expression was evaluated.

Dataset for analysis

The dataset used for the genome wide bioinformatics analysis was downloaded from the NCBI-SRA database. Whole genome sequence read of irradiated GM12878 (10 Gy, 4 h recovery; SRP022845) and CAL51 (5 Gy, 1 h recovery; ERP004219) cells and unirradiated controls were aligned and compared for missing reads.

Alignment, mapping, and visualization of the samples

The reads of the samples were aligned and mapped to the human hg38 reference genome using the alignment tool BOWTIE2, to generate an alignment file in the sam (SAM:

sequence alignment/map) format. Using SAMtools, the sam files were converted to their equivalent binary formats (BAM: binary alignment/map), which is required to upload and visualize the reads in genome viewer. For visualization, the Integrated Genome Viewer (IGV) was used where the samples were loaded as tracks for a visual comparison.

Extracting the nucleotide sequences, GC content and G4 motif number calculation

To extract the damaged sequences present in a stretch of more than 1000 bases, we first found the number of reads passing through every single nucleotide base position in the genome using a tool known as BEDtools suite. Stretches of more than 1000 bases with no reads passing through them were extracted using a AWK: a featured programming language in UNIX based operating systems. The locations of these stretches were extracted and their fasta sequences were obtained, using the BEDtools suite. To visualize the damaged stretches on the chromosome, 1% of these sequences were obtained from every chromosome through shell-scripting commands and represented on a karyogram using an R (scripting language) package called ggbio. The GC content of these sequences were calculated using AWK and the number of G4 motifs were obtained using the online tools Quadbase and Non-B DB.

Estimation of frequency of G4 motifs and occurrence of DNA breaks in different chromosomes

The percentage of DNA damage in every chromosome was determined as follows. The percentage of damaged regions = (the total length of all the damaged regions/ the length of the entire chromosome)*100. Using hclust (R package), the chromosomes were clustered in two distinct groups, viz. the chromosomes with < 15% damage and those with >15%. Analysis of G4 forming motifs in these groups revealed an Inverse correlation between extent of DNA damage and occurrence of G-quadruplex forming motifs, which was determined using Pearson's correlation coefficient. Subsequently, the average number of G4 motifs/damaged region was calculated. Percentage of damage was plotted on x-axis against the number of G4 motifs per damaged region on y-axis.

Distribution of G4s in the damaged and protected promoters and CDS for all the coding genes in the human genome was determined. The annotated human genome (hg38) protein coding gene locations (CDS in the genome) were mapped to all the unprotected regions (stretches of 1000 or more base pairs with no reads) across the entire genome using the BEDtools suite. The number of such mapped regions was obtained, and the percentage of genes with unprotected CDS was calculated using the equation: % unprotected CDS = (number of unprotected regions in the CDS /total number of CDS in the genome)*100. The unprotected

CDS regions were analyzed for the number of G4 forming units. The total number of G4 forming units in the human genome was obtained from a previous study (Kudlicki, 2016) and subsequently the percentage of G4 forming units in the unprotected CDS regions was calculated as follows: % of G4 forming units in the unprotected regions = (number of G4 forming units in the unprotected CDS / total number of G4 forming units in the CDS)*100. A similar pipeline was followed to obtain % of sequences unprotected in the promoter regions and the % of G4 forming units in the promoters.

Statistical analysis

Statistical significance was determined by GraphPad Prism 5 using the Student's *t* test or one way ANOVA test and the obtained values were considered significant if the p value was less than 0.05. Values are expressed as mean ± SEM. Statistical significance was determined by the GraphPad0-QuickCalcs software using the Student's t-test and Mann-Whitney U test and the values obtained were considered significant if the p value was less than 0.05. Besides, Pearson correlation studies were carried out and if the Correlation coefficient value R was less than -0.5, it was considered to be a negative correlation.

SUPPLEMENTARY REFERENCES

Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013). Quantitative visualization of DNA G-quadruplex structures in human cells. Nat Chem *5*, 182-186.

Bolte, S., and Cordelieres, F.P. (2006). A guided tour into subcellular colocalization analysis in light microscopy. J Microsc *224*, 213-232.

Das, K., Srivastava, M., and Raghavan, S.C. (2016). GNG Motifs Can Replace a GGG Stretch during G-Quadruplex Formation in a Context Dependent Manner. PloS one *11*, e0158794.

Dimitrova, N., Chen, Y.C., Spector, D.L., and de Lange, T. (2008). 53BP1 promotes non-homologous end joining of telomeres by increasing chromatin mobility. Nature *456*, 524-528.

Iyer, D., Vartak, S.V., Mishra, A., Goldsmith, G., Kumar, S., Srivastava, M., Hegde, M., Gopalakrishnan, V., Glenn, M., Velusamy, M*., et al.* (2016). Identification of a novel BCL2-specific inhibitor that binds predominantly to the BH1 domain. The FEBS journal *283*, 3408-3437.

John, F., George, J., Vartak, S.V., Srivastava, M., Hassan, P.A., Aswal, V.K., Karki, S.S., and Raghavan, S.C. (2015). Enhanced efficacy of pluronic copolymer micelle encapsulated SCR7 against cancer cell proliferation. Macromol Biosci *15*, 521-534.

Kudlicki, A.S. (2016). G-Quadruplexes Involving Both Strands of Genomic DNA Are Highly Abundant and Colocalize with Functional Sites in the Human Genome. PloS one *11*, e0146174.

Kumar, T.S., Kari, V., Choudhary, B., Nambiar, M., Akila, T.S., and Raghavan, S.C. (2010). Anti-apoptotic protein BCL2 down-regulates DNA end joining in cancer cells. The Journal of biological chemistry *285*, 32657-32670.

Nambiar, M., Goldsmith, G., Moorthy, B.T., Lieber, M.R., Joshi, M.V., Choudhary, B., Hosur, R.V., and Raghavan, S.C. (2011). Formation of a G-quadruplex at the BCL2 major breakpoint region of the t(14;18) translocation in follicular lymphoma. Nucleic acids research *39*, 936-948.

Nambiar, M., and Raghavan, S.C. (2010). Prevalence and analysis of t(14;18) and t(11;14) chromosomal translocations in healthy Indian population. Annals of hematology *89*, 35-43.

Nambiar, M., and Raghavan, S.C. (2012). Mechanism of fragility at BCL2 gene minor breakpoint cluster region during t(14;18) chromosomal translocation. The Journal of biological chemistry *287*, 8688-8701.
Nambiar, M., Srivastava, M., Gopalakrishnan, V., Sankaran, S.K., and Raghavan, S.C. (2013). G-quadruplex structures formed at the HOX11 breakpoint region contribute to its fragility during t(10;14) translocation in T-cell leukemia. Molecular and cellular biology *33*, 4266-4281.
Raghavan, S.C., Chastain, P., Lee, J.S., Hegde, B.G., Houston, S., Langen, R., Hsieh, C.L., Haworth, I.S., and Lieber, M.R. (2005). Evidence for a triplex DNA conformation at the bcl-2 major breakpoint region of the t(14;18) translocation. The Journal of biological chemistry *280*, 22749-22760.
Raghavan, S.C., Swanson, P.C., Wu, X., Hsieh, C.L., and Lieber, M.R. (2004). A non-B-DNA structure at the Bcl-2 major breakpoint region is cleaved by the RAG complex. Nature *428*, 88-93.
Rustighi, A., Tessari, M.A., Vascotto, F., Sgarra, R., Giancotti, V., and Manfioletti, G. (2002). A polypyrimidine/polypurine tract within the Hmga2 minimal promoter: a common feature of many growth-related genes. Biochemistry *41*, 1229-1240.
Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual, 2 edn (Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press).
Schiavone, D., Guilbaud, G., Murat, P., Papadopoulou, C., Sarkies, P., Prioleau, M.N., Balasubramanian, S., and Sale, J.E. (2014). Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. The EMBO journal *33*, 2507-2520.
Sebastian, R., and Raghavan, S.C. (2016). Induction of DNA damage and erroneous repair can explain genomic instability caused by endosulfan. Carcinogenesis *37*, 929-940.
Srivastava, M., Nambiar, M., Sharma, S., Karki, S.S., Goldsmith, G., Hegde, M., Kumar, S., Pandey, M., Singh, R.K., Ray, P.*, et al.* (2012). An inhibitor of nonhomologous end-joining abrogates double-strand break repair and impedes cancer progression. Cell *151*, 1474-1487.
Swanson, C., Saintigny, Y., Emond, M.J., and Monnat, R.J., Jr. (2004). The Werner syndrome protein has separable recombination and survival functions. DNA Repair (Amst) *3*, 475-482.
Wu, Q., Gaddis, S.S., MacLeod, M.C., Walborg, E.F., Thames, H.D., DiGiovanni, J., and Vasquez, K.M. (2007). High-affinity triplex-forming oligonucleotide target sequences in mammalian genomes. Molecular carcinogenesis *46*, 15-23.