SUPPLEMENTARY INFORMATION

Small molecule-induced DNA damage identifies alternative DNA structures in human genes

Raphaël Rodriguez[1,2,†], Kyle M. Miller[2,3,4,†], Josep V. Forment[2,3], Charles R. Bradshaw[2], Mehran Nikan[1], Sébastien Britton[2,3], Tobias Oelschlaegel[2,3], Blerta Xhemalce[2,4], Shankar Balasubramanian[1,5,6]* & Stephen P. Jackson[2,3]*

[1]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK.

[2]The Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QN, UK.

[3]Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge, CB2 1GA, UK.

[4]Department of Molecular Genetics and Microbiology, Institute for Molecular and Cellular Biology, University of Texas at Austin, 1 University Station A5000, Austin, TX 78712 USA.

[5]Cancer Research UK, Cambridge Research Institute, Li Ka Shing Center, Cambridge, CB2 0RE, UK.

[6]School of Clinical Medicine, University of Cambridge, Cambridge, CB2 0SP, UK.

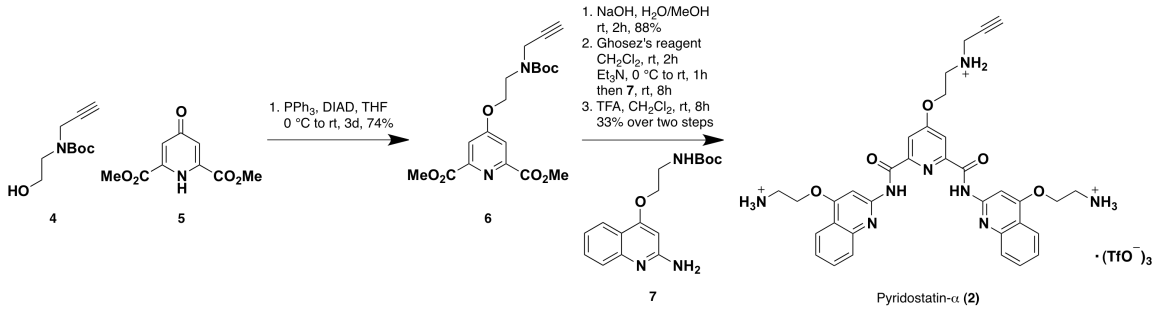
[†]These authors contributed equally to this work.

*Correspondence should be addressed to sb10031@cam.ac.uk or s.jackson@gurdon.cam.ac.uk.

SUPPLEMENTARY METHODS

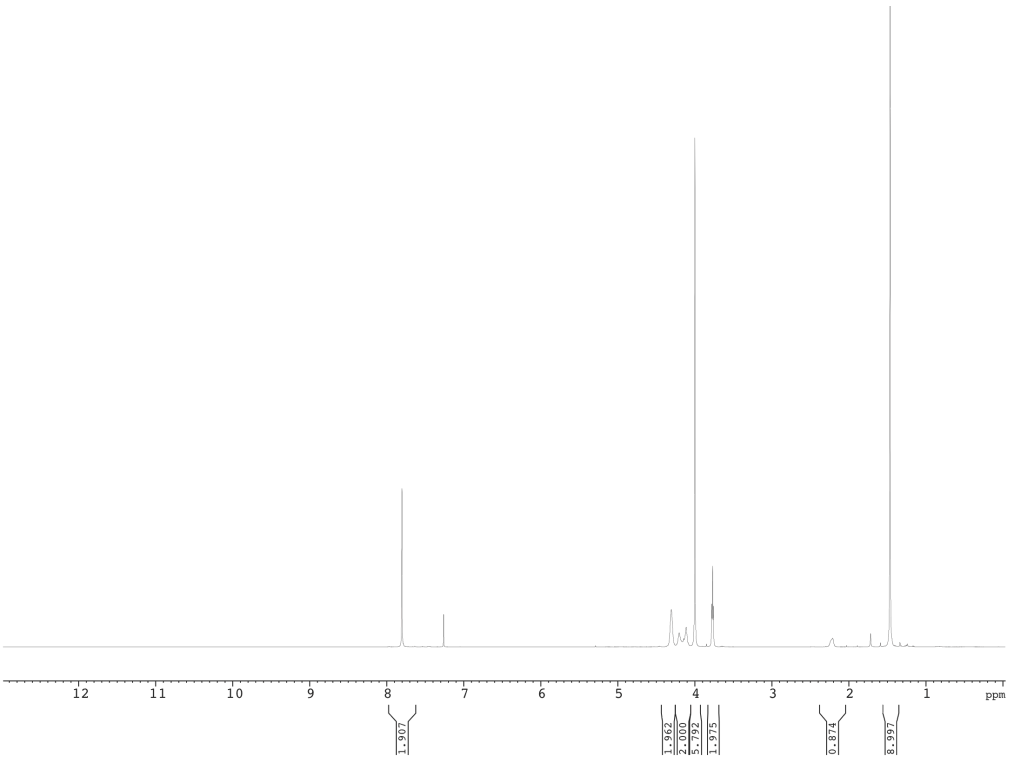
Synthetic procedures for pyridostatin-α (2)

All solvents and reagents were purified by standard techniques or used as supplied from commercial sources (Sigma-Aldrich). NMR spectra were acquired on Bruker DRX-400, DPX-400, DRX-500 instruments using deuterated solvents as detailed at 300 K. Notation for the ^{1}H NMR spectral splitting patterns includes: singlet (*s*), doublet (*d*), triplet (*t*), broad (*br*) and multiplet/overlapping peaks (*m*). Signals are quoted as δ values in ppm and coupling constants (*J*) are quoted in Hertz. Data analysis for NMR was performed using TopSpin® software. Mass spectra were recorded on a Micromass® Q-Tof (ESI) spectrometer. HPLC purification was carried out by using a Varian Pursuit C18, 5μ column (250 × 21.2 mm) and a gradient elution with H_2O/MeCN containing 0.1% TFA at a flow rate of 12.0 ml/min. Compounds **4**[51], **5**[17] and **7**[17] were synthesized according to previously established procedures. **1** was synthesized as previously described[17] and stored as a 10 mM stock solution of in water.

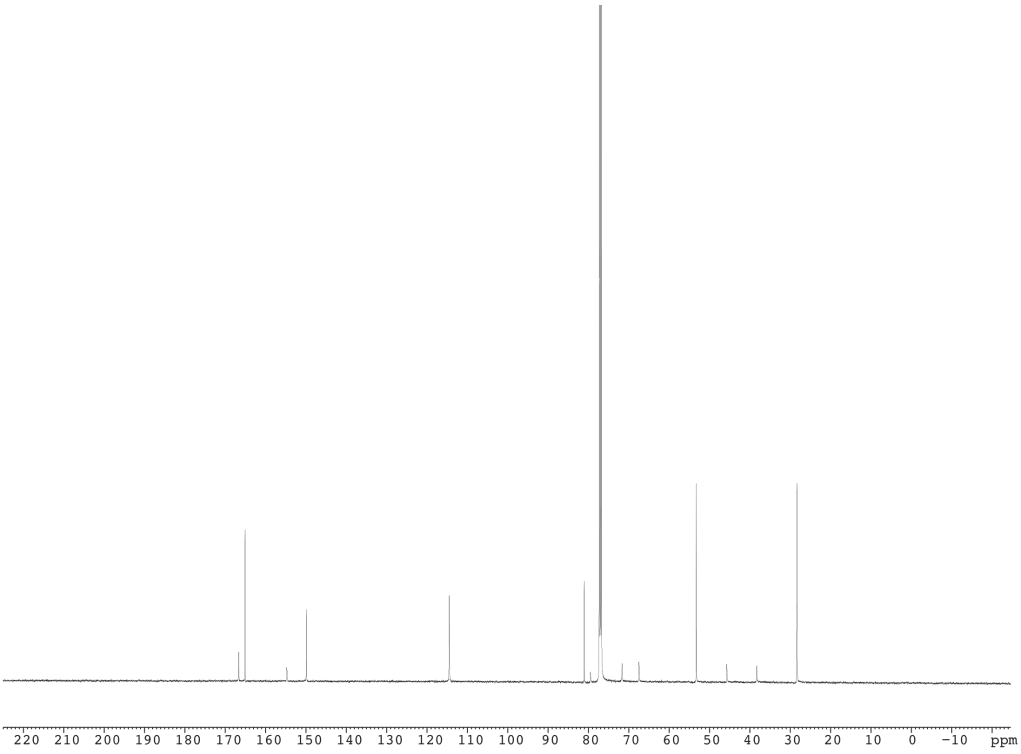


Compound 6

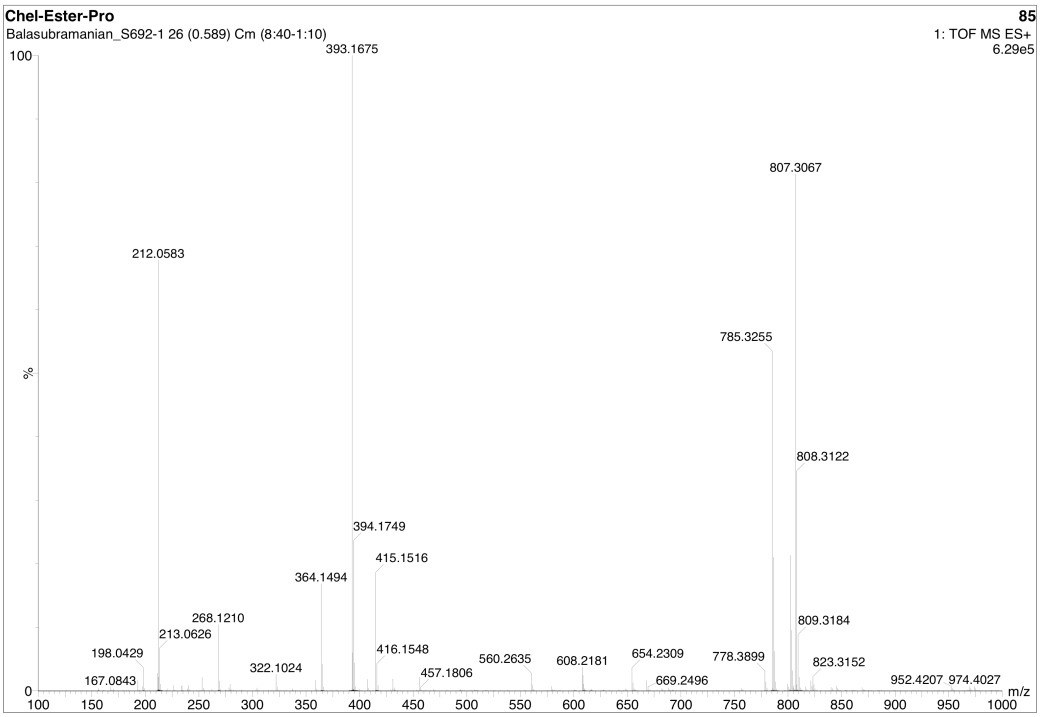
PPh_3 (1.12 g, 4.27 mmol), **4** (0.64 g, 3.21 mmol) and **5** (0.45 g, 2.13 mmol) were suspended in dry THF (15 ml) and stirred at 0 °C under argon. DIAD (0.57 ml, 2.89 mmol) was added dropwise and the mixture was stirred at rt for 3d. The solvent was evaporated under reduced pressure and the crude product purified by silica gel column chromatography (EtOAc-Petroleum Ether 1:1) to afford the product as a white solid (0.62 g, 74%). ^{1}H NMR (400 MHz, $CDCl_3$) δ_H 7.81 (*s*, 2H), 4.37-4.28 (*m*, 2H), 4.24-4.10 (*m*, 2H), 4.01 (*s*, 6H), 3.78 (*t*, *J* = 5.5 Hz, 2H), 2.25 (*br s*, 1H), 1.48 (*s*, 9H); ^{13}C NMR (100 MHz, $CDCl_3$) δ_c 166.7, 165.2, 154.8, 149.9, 114.6, 81.1, 79.6, 71.7, 67.6, 53.4, 45.8, 38.4, 28.4; HRMS (ES) calculated for $C_{19}H_{24}N_2O_7.H^+$: 393.1656, found 393.1675.



Compound **6** ^{1}H NMR spectrum



Compound **6** ^{13}C NMR spectrum

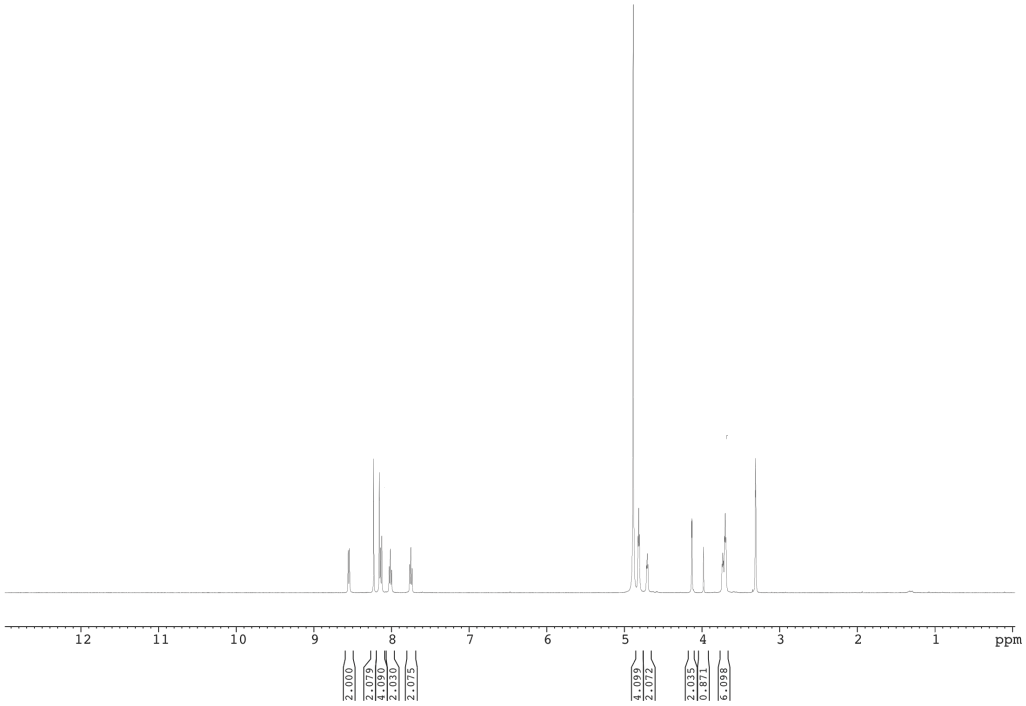


Compound **6** high-resolution mass spectrum

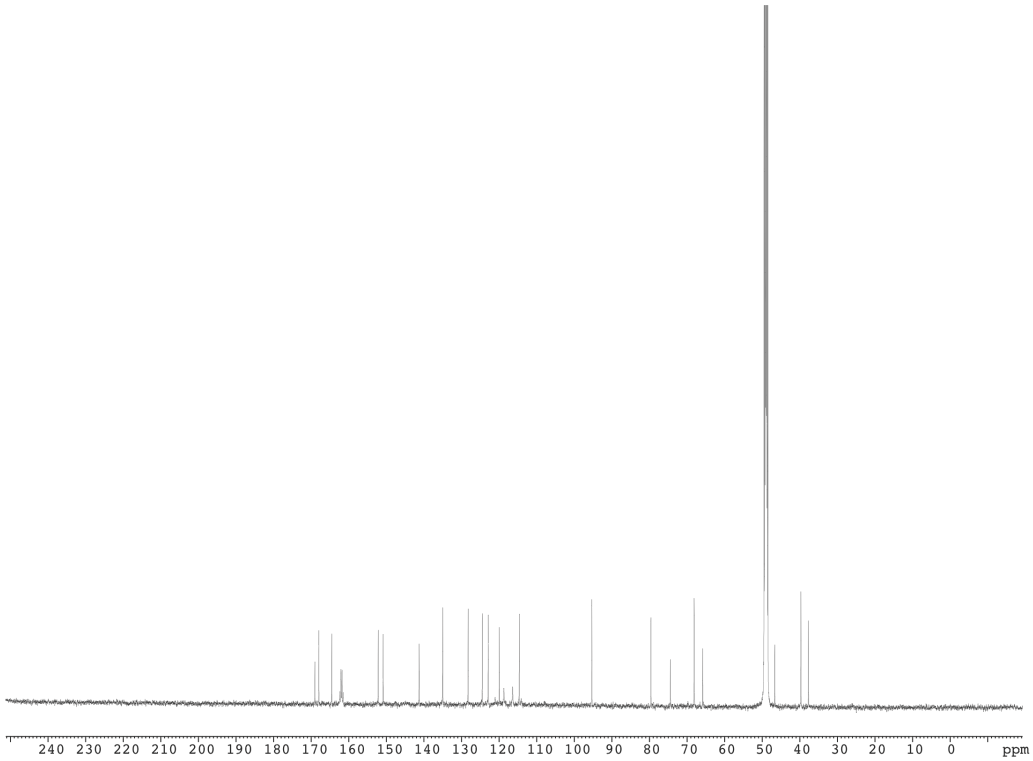
Pyridostatin-α (2)

Compound **6** (0.11 g, 0.28 mmol) was deprotected with NaOH (25 mg, 0.62 mmol) in 10 ml of H_2O/MeOH (1:1) for 2h. A 5% formic acid solution was added to reach pH 3.5. The mixture was then extracted with EtOAc, dried over $MgSO_4$ and evaporated to dryness. The crude product was used in the next step without further purification (0.09 g, 88%). 1-Chloro-*N*,*N*,2-trimethylpropenyl-amine (0.07 ml, 0.53 mmol) was added slowly to a solution of freshly deprotected compound **6** (0.09 g of di-acid, 0.25 mmol) in dry CH_2Cl_2 (2 ml) and the solution was stirred at rt under argon atmosphere for 2h. Dry triethylamine (0.07 ml, 0.50 mmol) was slowly added at 0 °C and the mixture stirred at rt for 1h. Compound **7** (0.16 g, 0.53 mmol) was added and the solution stirred at rt for 8h. The solvent was evaporated under reduced pressure and the crude was precipitated from hot acetonitrile to yield 0.08 g of a white solid. The crude product was stirred in TFA-CH_2Cl_2 (1:1) for 8h at rt. The mixture was concentrated to dryness and purified by HPLC to afford 0.08 g of **2** as a TFA salt (33% over two steps). Gradient: 10% MeCN in H_2O to 100% MeCN over 30 min (RT = 13 min). ^{1}H NMR (500 MHz, CD_3OD) δ_H 8.56 (*d*, *J* =

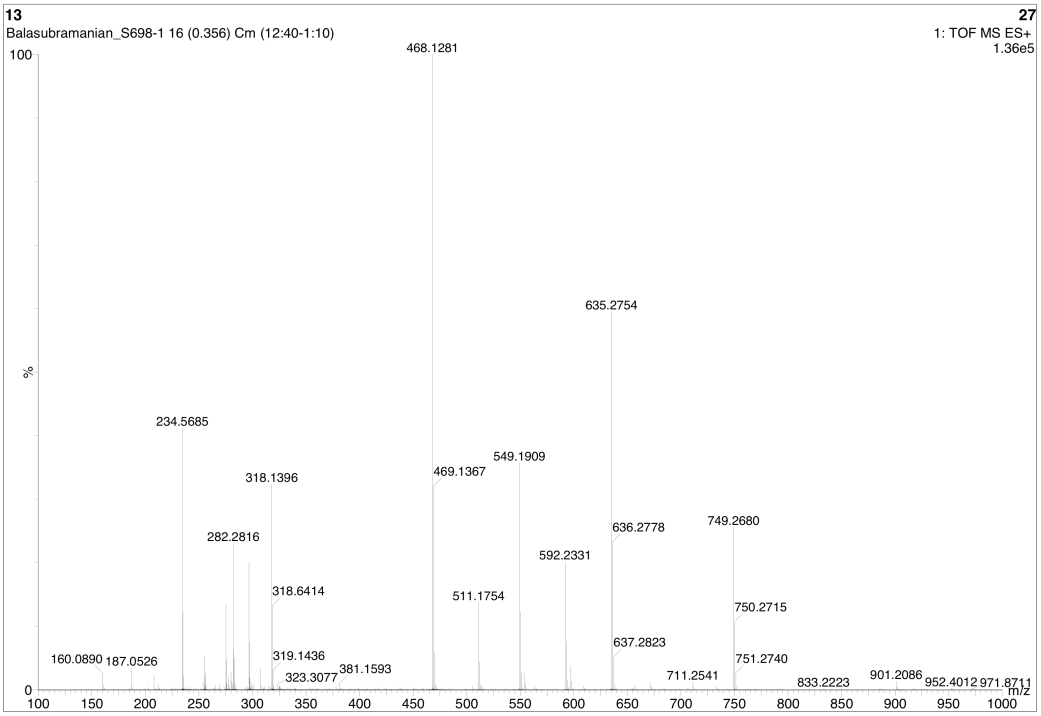
8.0 Hz, 2H), 8.22 (*s*, 2H), 8.13 (*d*, *J* = 8.5 Hz, 2H), 8.11 (*s*, 2H), 8.01 (*dd*, *J* = 8.5, 7.5 Hz, 2H), 7.76 (*dd*, *J* = 8.0, 7.5 Hz, 2H), 4.82 (*t*, *J* = 4.5 Hz, 4H), 4.70 (*t*, *J* = 4.5 Hz, 2H), 4.15 (*d*, *J* = 2.0 Hz, 2H), 3.97 (*t*, *J* = 2.0 Hz,1H), 3.73 (*br t*, *J* = 4.5 Hz, 2H), 3.68 (*t*, *J* = 4.5 Hz, 4H); ^{13}C NMR (125 MHz, CD_3OD) δ_c 169.0, 167.9, 164.5, 152.1, 150.9, 141.3, 135.0, 128.2, 124.4, 122.8, 119.9, 114.6, 95.3, 79.6, 74.4, 68.1, 65.8, 46.6, 39.7, 37.7; HRMS (ES) calculated for $C_{34}H_{34}N_8O_5.H^+$: 635.2725, found 635.2754.



2 ^{1}H NMR spectrum



2 ^{13}C NMR spectrum



2 high-resolution mass spectrum

Cell culture, reagents and treatments

Human MRC5-SV40, MO59K[52], MO59J[52], U2OS and MDA-MB-231 cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 100 U/ml penicillin, 100 mg/ml streptomycin and 2 mM L-glutamine. For cell cycle analysis, cells were fixed in ice-cold 70% ethanol, stained with propidium iodide (10 mg/ml PI, 250 mg/ml RNase A, 0.5% BSA, 0.02% sodium azide in 1X PBS) and analyzed by FACS on a FACS Calibur using standard settings. Pyridostatin was used at 2 μM unless stated otherwise. For inhibitors, ATMi (20 μM, KU-55933, obtained from Tocris), DNA-PKi (2 μM, NU7441, obtained from Tocris), Caffeine (4 mM) and Chk1/Chk2i (50 nM, AZD7762, obtained from AstraZeneca) were added to cells 1 h prior to pyridostatin treatment. PARP inhibitor (10 μM, KU-0058948) was added 1 h prior to addition of **1**. The transcription inhibitor DRB was used at 100 μM and added to cells 1 h prior to addition of **1**. To inhibit replication, aphidicolin (5 μM) was added 2 h pre-pyridostatin treatment. All inhibitors remained in cells for the duration of the experiment. The topoisomerase poison doxorubicin (Dox, Sigma) was used at a final concentration of 100 nM for the indicated times.

Protein extracts and western blotting

For whole cell extracts, cells were washed once with PBS (phosphate buffered saline), collected by adding Laemmli buffer (4% (v/v) SDS, 20% (v/v) glycerol and 120 mM Tris-HCl pH 6.8), boiled for 5 min at 95 °C, sheared through a 23-gauge needle and boiled again before loading. Samples were resolved by SDS-PAGE and analyzed by standard western blotting techniques. SRC quantification was performed using a LI-COR Odyssey infrared imaging system (LI-COR Biosciences) where densitometry values were calculated for a loading control (H2AX) and SRC values were normalized to control values. Secondary antibody used for quantification is IRDye 680CW Donkey anti-rabbit (LI-COR Biosciences). For standard techniques, antigens were detected by chemiluminescence (ECL; Amersham). Secondary antibodies used for ECL were goat anti-rabbit HRP (Perbio Science Ltd) and rabbit anti-mouse HRP (Dako Ltd). Primary antibodies used for western blotting in this study are γH2AX (Upstate), H2AX (Abcam), RPA pS4/8 (Bethyl Laboratories, Inc), RPA (Abcam), KAP1 pS824 (Bethyl

Laboratories, Inc), KAP1 (Santa Cruz Biotechnology, Inc), Chk1 pS345 (Cell Signaling), Chk1 (Santa Cruz Biotechnology, Inc), p21 (Santa Cruz Biotechnology, Inc), PARP-1 (Cell Signaling), DNA-PK pS2056 (Abcam), DNA-PK (Abcam), Tubulin (Sigma), SRC (Abcam).

Immunofluorescence analyses

For IF analysis of mitotic spreads, MRC5-SV40 cells were untreated or treated with 2 μM **1** for 24 h in 6 cm dishes. Cells were treated with 0.1 μg/ml colcemid (Gibco) for 1 h followed by incubation with Chk1/Chk2i for 2 h to abrogate the G2/M checkpoint. To collect cells, culture medium was pipetted off and placed in a 15 ml conical tube. Cells were washed once with PBS followed by trypsinization. 10 ml of medium were added to each 6 cm dish and cells were collected by gentle pipetting. This medium was then added to the first 4 ml that were collected. Cells were pelleted by centrifugation at 1,000 rpm for 5 min and all but 300 μl of the supernatant were removed. Cells were gently re-suspended in the remaining medium and kept on ice. 1 ml of ice-cold hypotonic 0.075 M KCl was added dropwise to the suspension with gentle agitation using a vortexer on a low speed. An additional 2-3 ml of KCl were added and cells were incubated at 37 °C for 15 min to allow swelling of the cells. Cells were centrifuged onto coated Shandon Cytoslides (Thermo Scientific) at 1000 rpm for 5 min using a Cytospin. Slides were pre-incubated in KCM buffer (120 mM KCl, 20 mM NaCl, 10 mM Tris-HCl pH 7.2, 0.5 mM EDTA pH 8, 0.1% (v/v) Triton X-100, and protease inhibitors) for 30 min. Slides were blocked for 30 min in KCM containing 1% BSA followed by 1 h incubation in mouse anti-γH2AX (Upstate) primary antibody. Slides were washed 3x with KCM buffer followed by 45 min incubation in the mouse anti-rabbit Alexa Fluor 488 secondary antibody. Slides were washed 3x with KCM buffer and fixed in 2% PFA in KCM buffer for 20 min at rt followed by 3x washes with KCM. Coverslips with Vectashield containing DAPI (Vector laboratories) were fixed over the cells. Samples were imaged with an inverted FV1000 confocal microscope (Olympus).

Neutral comet assays

MRC5-SV40 cells were grown in 6 cm dishes either untreated or treated with 2 μM **1** for 24 h. DNA-PKcs inhibitor was administered as described above. Cells were washed once with Ca^{2+}/Mg^{2+} free PBS (Invitrogen) then scraped into 5 ml of Ca^{2+}/Mg^{2+} free PBS. Cells were pellet by centrifugation at 1000 rpm for 5 min at 4 °C. Cells were re-suspended in 500 μl of PBS and kept on ice. Cells were then processed using the comet assay kit from Trevigen according to the manufacturer's protocol. Briefly, cells were mixed with melted agarose (Trevigen LMAgarose) at a 1/10 cell/agarose volume ratio, mixed thoroughly by pipetting up and down several times and then 75 μl solution containing cells in agarose was placed on gelbond film (0.22 mm thick, LONZA agarose gel support medium). A 22 mm glass coverslip was added immediately and gently pressed until all the agarose was evenly spread. Slides were placed at 4 °C for at least 30 min. Glass coverslips were removed and slides were incubated for 30 min in lysis solution that was pre-chilled on ice. Slides were washed with 50 ml of 1X TBE for 5 min at rt. Slides where then electrophoresed for 10 min at 1V/cm in 1X TBE. Slides were washed briefly in water and immersed in 70% ethanol for 5 min. Slides were air dried at rt overnight followed by staining with SYBR green (Invitrogen), in TE solution (10 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8.0) for 5 min at 4 °C followed by 1 h of drying at rt. Images were collected using a GFP filter on an Olympus epifluorescence microscope. Comets were analyzed from at least 100 cells using Cometscore software (TriTek) to determine tail moment. Data from three independent experiments were collected and plotted using Prism software. Error bars represent the S.E.M.

FRET-melting assay

The following dual labelled oligonucleotides were purchased from Eurogentec® Ltd. H-telo (5′-FAM-GGG TTA GGG TTA GGG TTA GGG-TAMRA-3′)[18], C-kit1 (5′-FAM-GGG AGG GCG CTG GGA GGA GGG-TAMRA-3′)[18], C-kit2 (5′-FAM-GGG CGG GCG CGA GGG AGG GG-TAMRA-3′)[18], C-myc (5′-FAM-TGA GGG TGG GTA GGG TGG GTA A-TAMRA-3′)[18], KRAS (5′-FAM AGG GCG GTG TGG GAA GAG GGA AGA GGG GGA GG-TAMRA-3′)[53] and ds-DNA (5′-FAM-TAT AGC TAT A-HEG-T ATA GCT ATA-TAMRA-3′)[18]. Oligonucleotides were pre-annealed in potassium

cacodylate buffer (60 mM, pH 7.4) at 400 nM by heating at 94 ºC for 10 min followed by slow cooling to room temperature at a controlled rate of 0.1 ºC/min. 96-well plates were prepared by addition of 50 μl of the annealed DNA solution to each well, followed by 50 μl solution of **1** or **2** at the appropriate concentration. Measurements were made in triplicate with an excitation wavelength of 483 nm and a detection wavelength of 533 nm. Analysis was carried out using OriginPro 7.5 data analysis and graphing software (OriginLab®). Fluorescent readings for FRET experiments were performed using a Roche® LightCycler™ _480 System RT-PCR.

Circular dichroism spectroscopy

DNA oligonucleotides were purchased from Eurogentec® Ltd. CD spectra were recorded on an Applied Photophysics Chirascan circular dichroism spectropolarimeter using a 1 mm path length quartz cuvette. Scans were performed at 25 °C over a wavelength range of 210-330 nm with response time of 0.5 s, 1 nm pitch and 1 nm bandwidth. Blank spectra of samples containing buffer were subtracted from DNA samples. The CD spectra represent an average of three scans, are zero-corrected at 330 nm and normalized (Molar ellipticity θ is quoted in 10^5 deg cm^2 dmol^{-1}). The sequences were used at 10 μM in the presence of buffer containing Tris-HCl (10 mM, pH 7.4) and 100 mM of KCl. Samples were pre-annealed by heating at 95 °C for 5 min and then cooled to rt over 6-8 h.

Nuclear magnetic resonance spectroscopy

DNA oligonucleotides were purchased from Eurogentec® Ltd. and were HPLC purified. ^{1}H NMR spectra were recorded on a 500 MHz TCI-ATM Cryo instrument at 298 K unless otherwise stated. Titration with **1** was performed at 313 K and scan were recorded 15 mins after addition of **1**. Water suppression was achieved using excitation sculpting. Samples were prepared in H_2O containing phosphate buffer (20 mM, pH 7.4) and KCl (100 mM), and were pre-annealed by heating at 95 °C for 5 min and then cooled down to room temperature over 6-8 h. Samples were used at 100-200 μM in a 1:9 mixture of D_2O/H_2O at pH 7.4.

GFP and GFP-hPif1 expressing stable cell lines

cDNA encoding the human Pif1 alpha (hPif1α) isoform was amplified from IMAGE clone 9021504 and cloned between the XhoI and BamHI restriction sites of pEGFP-C1 plasmid. Individual clones stably expressing GFP or GFP-hPif1α were isolated after selection with 0.5 mg/ml G418 for two weeks following transfection with pEGFP-C1 or pEGFP-C1-hPif1α in U2OS cells. Cells from individual clones expressing either pEGFP-C1 or pEGFP-C1-hPif1α were CSK treated prior to 2% PAF fixation, and GFP signal was enhanced by using a rabbit anti-GFP antibody (Invitrogen) detected with Alexa Fluor 488 coupled anti-rabbit secondary antibody (Invitrogen). For imaging, 0.2 μm z-stacks were acquired using a Deltavision microscope (Applied Precision) with an 100X objective. Resulting stacks were deconvolved using softWoRx software. Pictures presented in the manuscript correspond to one deconvolved stack. Orthogonal views were used to confirm co-localization of **3** and GFP-hPif1α foci.

Chromatin Immunoprecipitation (ChIP) and qPCR

ChIP analyses were performed as previously described[48]. Briefly, untreated or pyridostatin treated MRC5-SV40 cells were fixed with 1% formaldehyde for 15 min at rt and quenched with 20 mM glycine for 5 min at rt. Cells were washed with PBS, and collected by scraping in cold PBS followed by a centrifugation for 30 min at 13000 rpm at 4 °C. The cell pellets were re-suspended in 2 ml of cold RIPA Buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 2 mM EDTA pH 8, 1% (v/v) NP-40, 0.5% (v/v) sodium deoxycholate, 0.1% (v/v) SDS) supplemented with protease inhibitors (Complete, EDTA-free from Roche). The suspension was sonicated in 15 ml conical tubes three times for 8 min at maximum setting (30 sec ON/OFF cycles) in a cooled Bioruptor® (Diagenode). 10 μg of chromatin were incubated with a γH2AX (Epitomics) antibody overnight at 4 °C, then with pre-washed Protein A+G Dynabeads® (Invitrogen) for 2 h at rt. Beads were washed for 5 min at rt with TSE-150 (0.1% (v/v) SDS, 1% (v/v) Triton X-100, 150 mM NaCl, 2 mM EDTA, 20 mM Tris-HCl pH 8), TSE-500 (0.1% (v/v) SDS, 1% (v/v) Triton X-100, 500 mM NaCl, 2 mM EDTA, 20 mM Tris-HCl pH 8), LiCl buffer (0.25 M LiCl, 1% (v/v) NP-40, 1% (v/v) DOC, 1 mM EDTA, 10 mM Tris-HCl pH 8) and TE (10 mM Tris-HCl pH 8, 0.1 mM EDTA). Immuno-complexes were eluted with elution buffer (100

mM sodium bicarbonate, 1% (v/v) SDS) for 30 min at 30 °C and de-crosslinked overnight at 65 °C. ChIPed samples were purified into water with the Qiaquick® PCR purification kit (Qiagen). To validate γH2AX induction upon pyridostatin treatment, qPCR was performed on samples obtained by ChIP analysis, which were performed as described above. From each γH2AX immunoprecipitation, the eluted sample was used in Real-time quantitative PCR analysis on an ABI PRISM 7000 /7300 sequence detection system with the use of SYBR® Green (Applied Biosystems) for each primer pair. The sequences of the primers used for the qPCR analysis are: *HRAS* forward (GGAGCCCTACTTCCAAGGTC) and reverse (TGAGAGGCATCACCTTTCCT), control gene *KRT80* forward (GGAGTAGGGCTTGAGCACAG) and reverse (GAGCAGATCCTCGGACACTC), *MYC* forward (CCTTAGCATTCGCTTTCCAG) and reverse (AGTCTGGGTGTGGGCATAAG), *SRC*_1 forward (CTTTGGCTCCGTTTCTGTGT) and reverse (CCACTTCCTAACGGGAACAA), and *SRC*_2 forward (CTCCCTGCAGACATGACCTT) and reverse (GTGGGAAGGGAGAAGAGGAC).

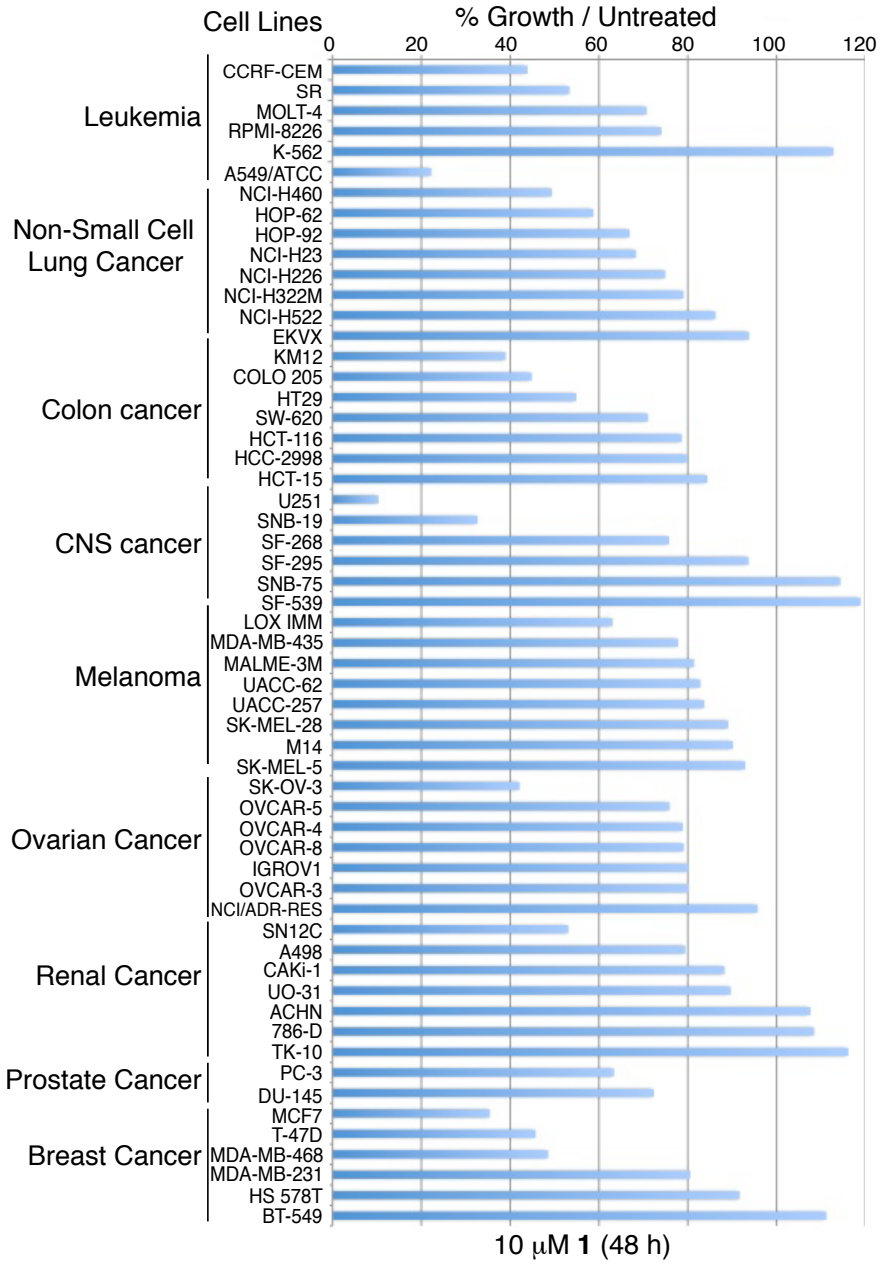
ChIP-Sequencing and bioinformatics

We analyzed all oncogenes and tumor suppressors as defined by CancerGenes, which contains 3164 cancer-associated genes comprising 385 designated oncogenes and 763 tumor suppressors[32]. ChIP-Seq libraries of γH2AX from untreated and pyridostatin treated cells were loaded into the UCSC browser and settings were manually set with a cut off at 5 reads. Genes from our selected set were scored for their enrichment of γH2AX in pyridostatin treated versus untreated samples, with the following scheme: no (no γH2AX signal), yes (mild γH2AX-enrichment occurred over a partial region of the gene), yes (+; moderate γH2AX-enrichment occurred over a partial region of the gene) and yes (++; γH2AX-enrichment occurred over the entire length of the gene). The yes (++) category is displayed in **Supplementary Fig. 10** and the scoring of all genes is shown in **Supplementary Dataset 2**.

RNA analysis

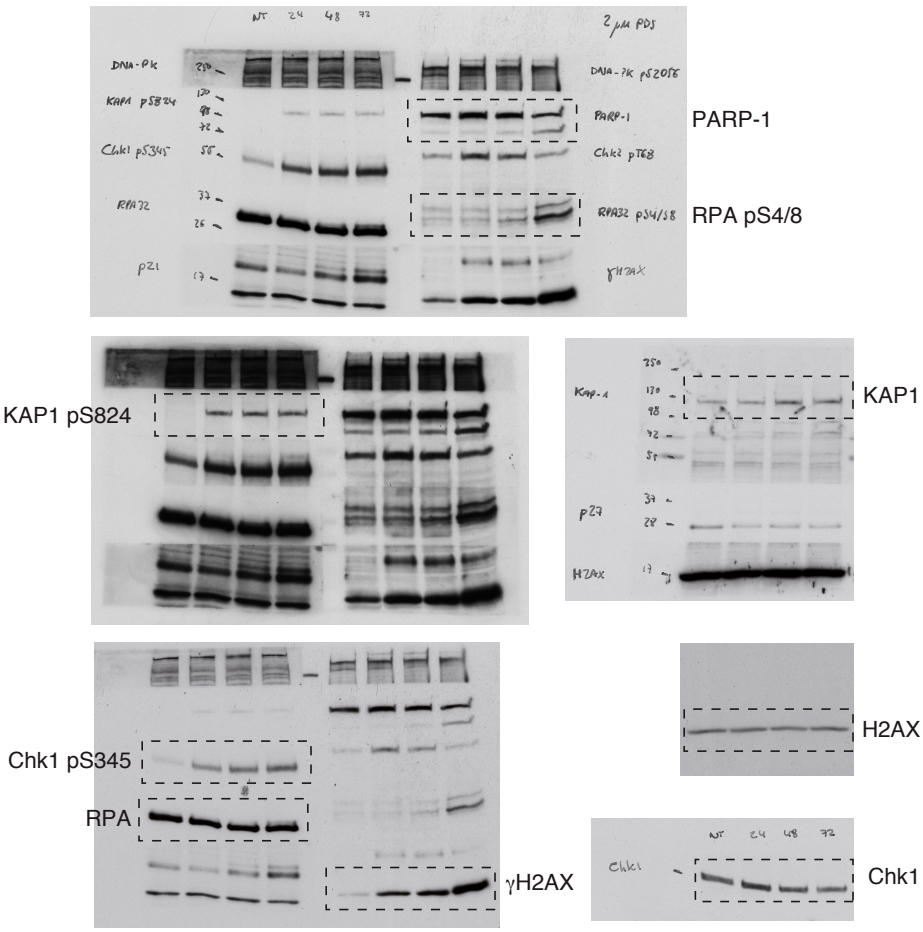
MRC5-SV40 or MDA-MB-231 cells were untreated or treated with 2 μM **1** or 100 nM doxorubicin (Dox) for the indicated times. Total RNA was purified from each sample using a RNA/protein purification kit from Norgen with Turbo DNase from Ambion following the manufacturer's protocol. cDNA synthesis of 500 ng of total RNA was performed with the Superscript III First-Strand Synthesis System (Invitrogen) using oligo(dT) primers. For quantitative Real-Time PCR (qRT-PCR) analyses, 1/50th of each reaction was used. The following gene specific Quantitect primer assays (Qiagen, cat. number included) were used in our analysis: *ALAS1* (QT0073122), *B2M* (QT00088935), *DDX1* (QT00065205), *DDX51* (QT00493143), *HRAS* (QT1668338), *RHOC* (QT01679412), *SRC* (QT00039326), *SREBF1* (QT00036897), *FAIM2* (QT00034125), *VAV2* (QT00002009), *MYC* (QT00035406), *CST6* (QT00023912), *NOTCH1* (QT01005109), *C9orf140* (QT000200970), *LLGL1* (QT00090020), *AHRR* (QT00010388). RT-PCR analysis was performed using SYBR green (Applied Biosystems) for detection with an Applied Biosystems StepOne Plus system.

SUPPLEMENTARY RESULTS

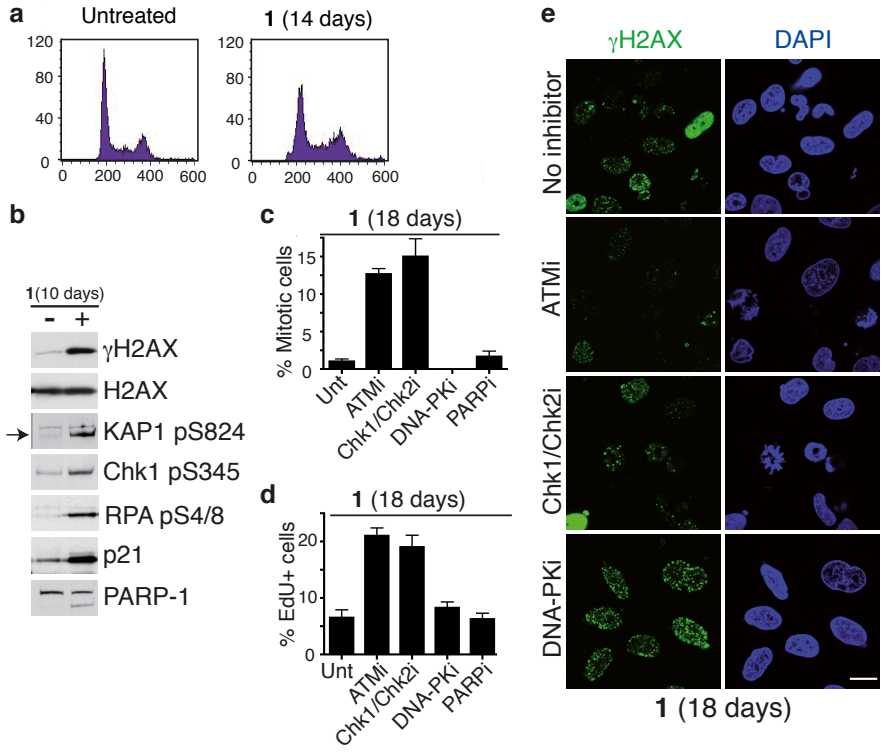


Supplementary Figure 1 | Pyridostatin inhibited growth of various cancer cell lines. **1** was screened using the In Vitro Cell Line Screening Project (IVCLSP) from the Developmental Therapeutics Program from the NCI/NIH. **1** was incubated for 48 h with over 60 different cancer cell lines at a concentration of 10 μM, and percentage cell growth was measured and compared to untreated samples. Analyses were done as described at http://dtp.cancer.gov/branches/btb/ivclsp.html.

Fig. 1d



Supplementary Figure 2 | Full gel images for blots shown in Fig. 1d. After electrophoresis and transfer, blots were cut into strips that contained the protein of interest and were stained with the indicated antibodies. Dashed boxes represent the regions displayed in **Fig 1d**. Note that H2AX and Chk1 were individually analyzed as single blots.



Supplementary Figure 3 | Long-term pyridostatin treatments caused sustained cell cycle arrest and DNA damage signaling. (**a**) FACS analysis of MRC5-SV40 cells treated with 2 μM **1** for 14 days; fresh medium and drug were added every two days for the duration of the experiment; cell counts (y-axis) and DNA content (x-axis) are indicated. (**b**) **1** induced DNA damage signalling pathways involved in DSB repair and p21 expression; cells were analyzed as in **Fig. 1d** except that **1** was present for 10 days. (**c**) G2/M arrest in cells treated with **1** was abrogated by ATM or Chk1/Chk2 inhibition; cells were treated for 18 days followed by incubation with the indicated inhibitors as described in **Methods**. Cells were fixed and stained with DAPI after 2 h of inhibitor treatment. Over 200 cells were counted and mitotic figures scored and plotted. (**d**) Inhibition of ATM (ATMi) or Chk1/Chk2 (Chk1/Chk2i) resumed replication in long-term pyridostatin treated cells. Experiments were performed as in **Supplementary Fig. 3c** except that the EdU protocol was performed as described in **Methods**; graphs represent three independent experiments; over 200 cells were counted for each condition and error bars represent S.E.M. (**e**) DNA-PK inhibition (DNA-PKi) enhanced γH2AX in long-term **1** treated cells; cells were analyzed as in **Supplementary Fig. 3c** except that following inhibitor treatment, cells were analyzed by IF with the indicated antibodies. MRC5-SV40 cells were used throughout all **Supplementary Results** and were either untreated or treated with 2 μM **1** during 24 h unless otherwise stated. Scale bar, 10 μM.

Fig. 2a

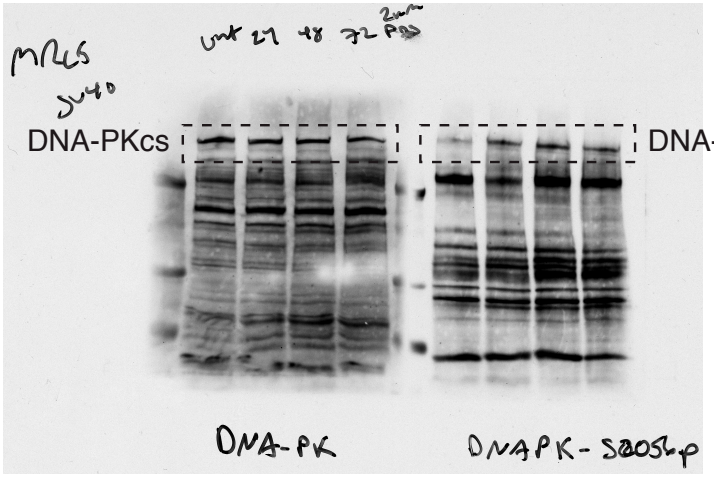
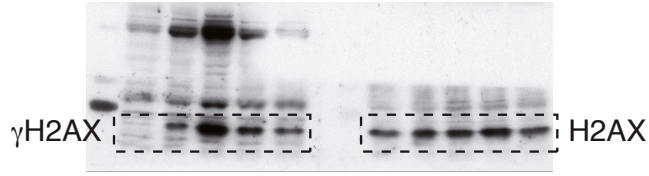
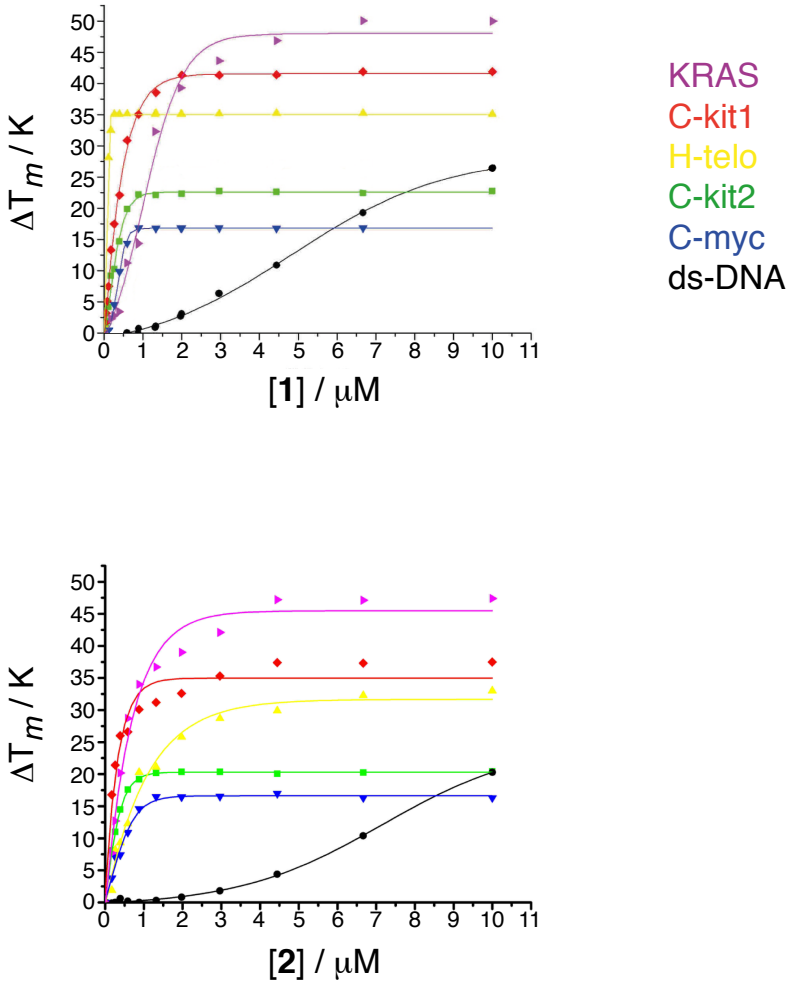


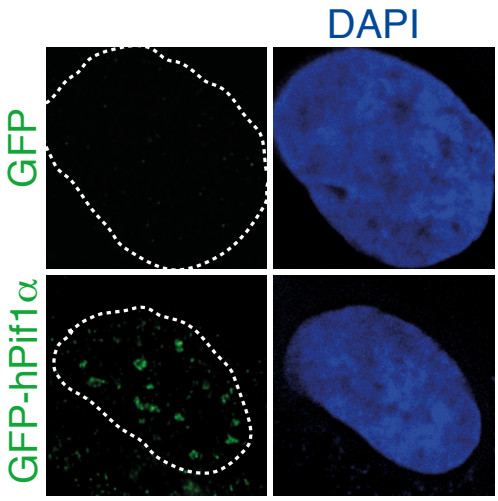
Fig. 2b



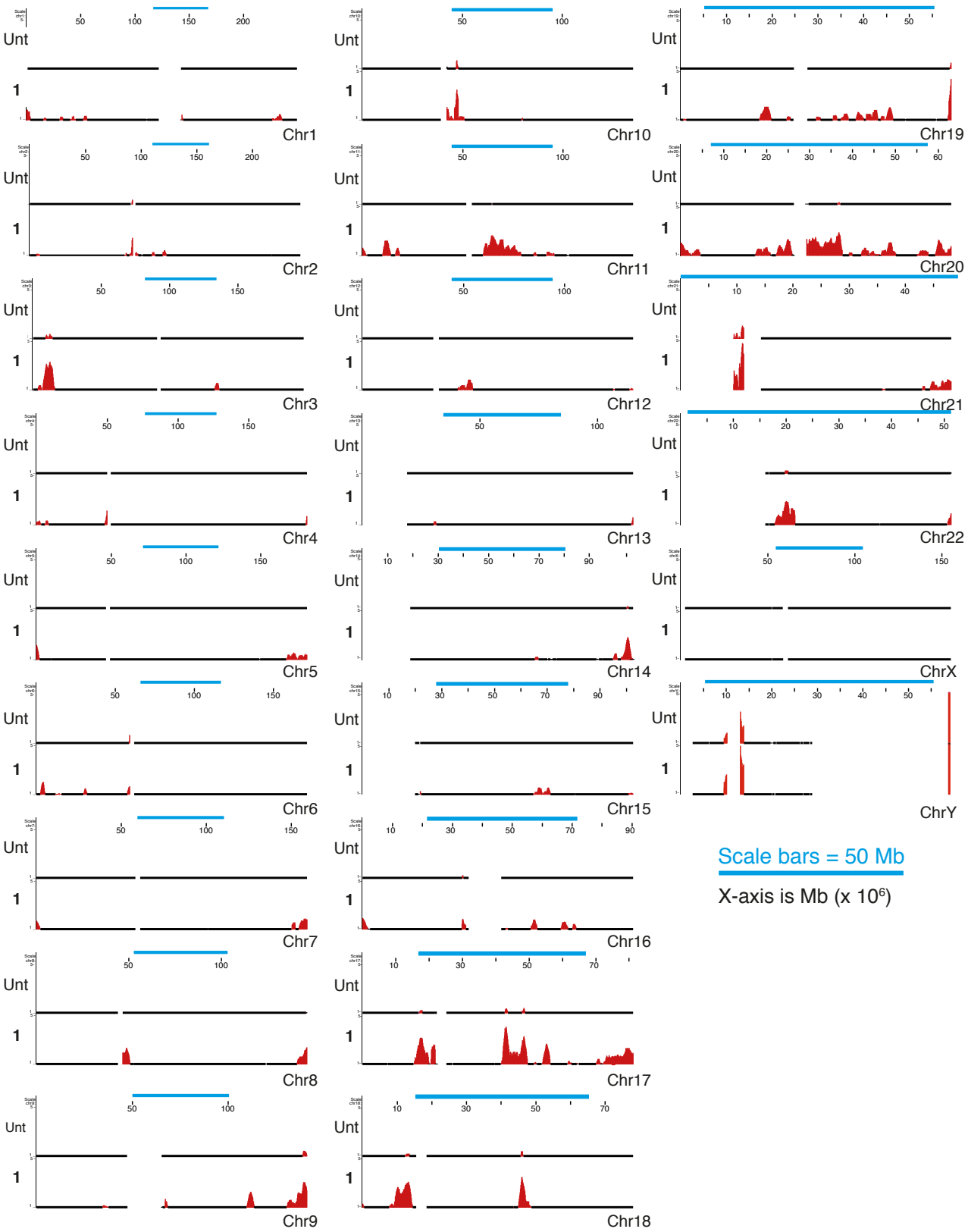
Supplementary Figure 4 | Full gel images for blots shown in Fig. 2a,b. After electrophoresis and transfer, blots were cut into strips that contained the protein of interest and were stained with the indicated antibodies. Dashed boxes represent the regions displayed in **Fig 2a,b**. Note that DNA-PKcs and DNA-PKcs pS2056 were individually analyzed as single blots.



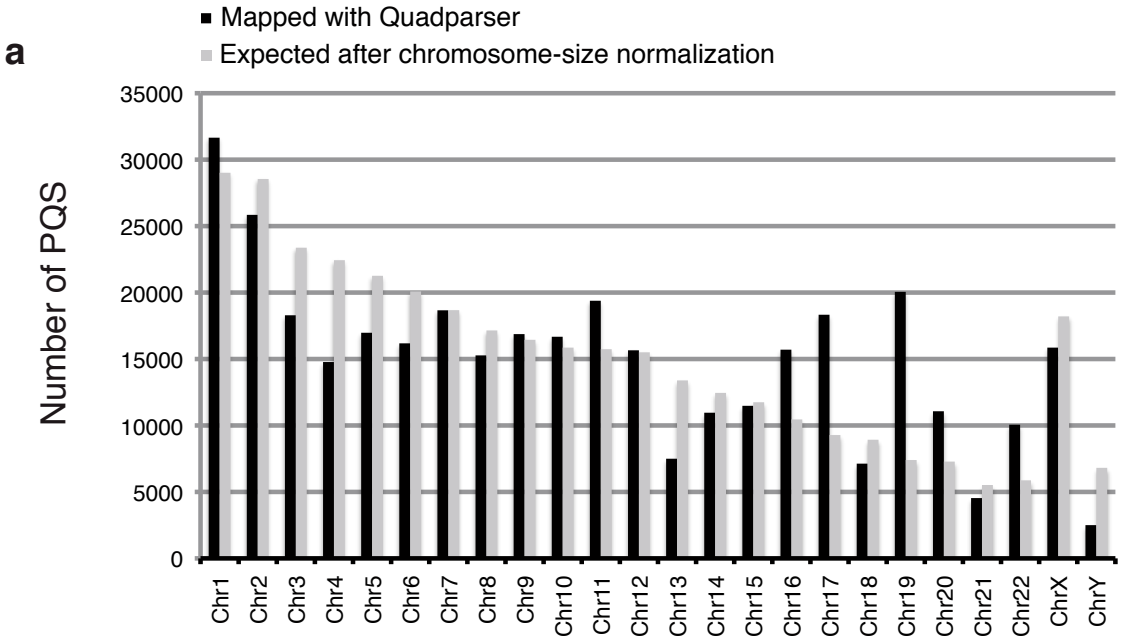
Supplementary Figure 5 | FRET-melting profiles of diverse G-quadruplex structures stabilized by pyridostatin (1) and pyridostatin-α (2). Distinct PQS were pre-annealed in potassium containing buffer and melted in the presence of increasing amounts of **1** or **2,** showing high stabilization potentials of the drugs regardless of the nature of loop sequences.

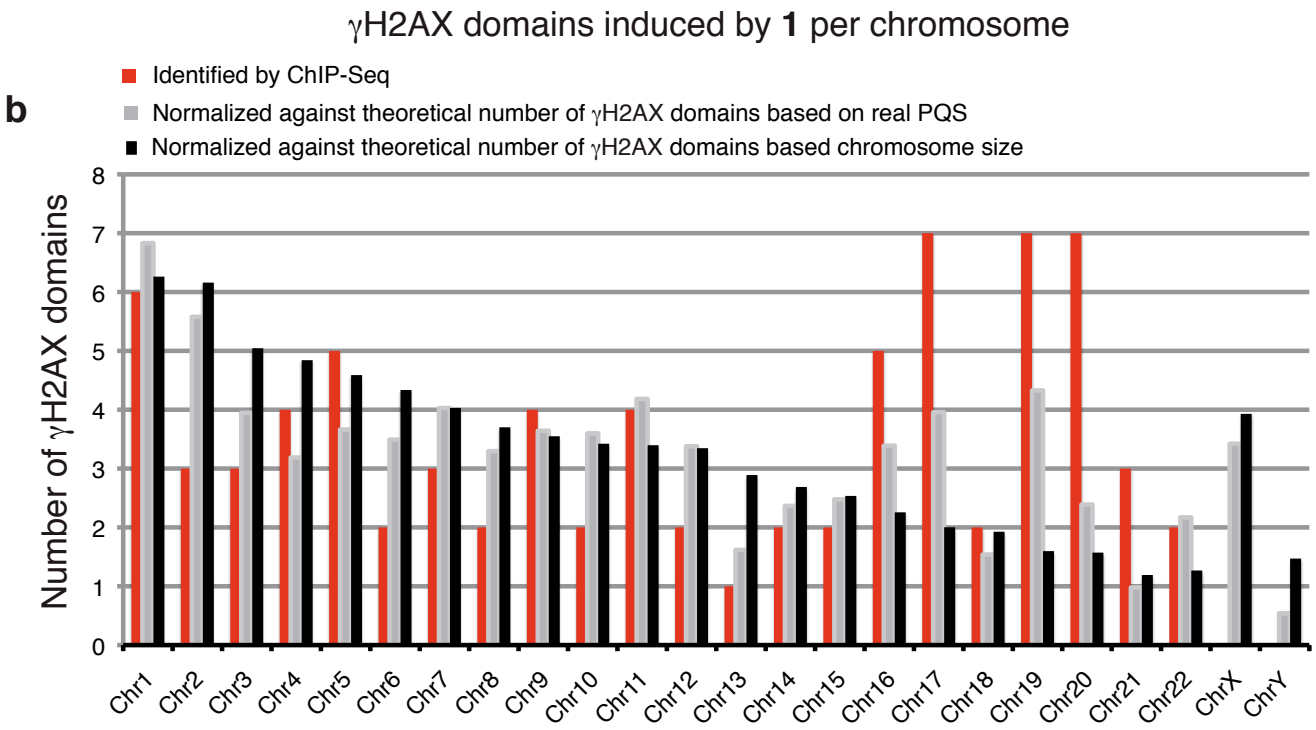


Supplementary Figure 6 | GFP control and GFP-hPif1α expressing U2OS cells. GFP-hPif1α forms small nuclear foci in absence of drug treatment. Note that cell were first pre-extracted with CSK buffer as described in **Supplementary Methods**, then fixed with formaldehyde and stained with the indicated antibody.

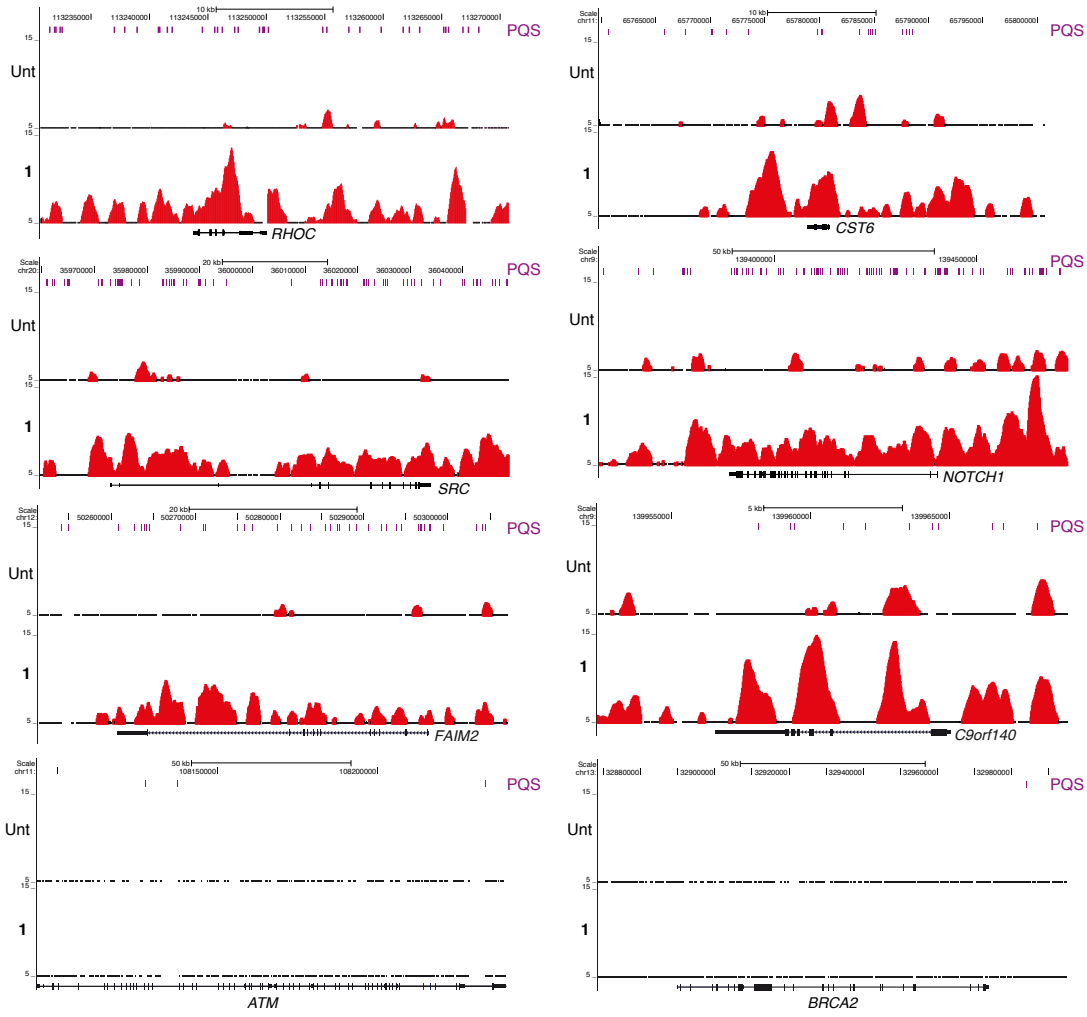

Unt
1
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr20
Chr21
Chr22
ChrX
ChrY
Scale bars = 50 Mb
X-axis is Mb (x 10⁶)

Supplementary Figure 7 | ChIP-Seq profiles of γH2AX in untreated compared to cells treated with pyridostatin. Figures represent each human chromosome; images were taken from the UCSC browser and files were labelled for clarity. For each chromosome, the γH2AX profile in untreated cells is above the γH2AX profile of cells treated with **1**; y-axis represents the enrichment, and γH2AX peaks have been highlighted in red, chromosome coordinates are given in megabases (Mb, x 10^6 bases) and scale bars (blue) equal 50 Mb. Profiles were obtained from uniquely mapping sequence reads, thereby excluding telomeres and rDNA.


a
Mapped with Quadparser
Expected after chromosome-size normalization
Number of PQS
35000
30000
25000
20000
15000
10000
5000
0
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr20
Chr21
Chr22
ChrX
ChrY


γH2AX domains induced by 1 per chromosome
b
Identified by ChIP-Seq
Normalized against theoretical number of γH2AX domains based on real PQS
Normalized against theoretical number of γH2AX domains based chromosome size
Number of γH2AX domains
8
7
6
5
4
3
2
1
0
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr20
Chr21
Chr22
ChrX
ChrY

Supplementary Figure 8 | γH2AX-enrichments compared to PQS at the chromosomal level in pyridostatin treated cells. (**a**) Some chromosomes were enriched in PQS; black histograms represent real numbers of PQS mapped from the Feb 2009 (GRCH37/hg19) human genome by using QuadParser (www.quadruplex.org/?view=quadparser, bed file available at http://bioinformatics.gurdon.cam.ac.uk/Rodriguez/Rodriguez_PQS_NCBI37.bed), grey histograms represent the expected number of PQS obtained after chromosome-size normalization. Note that chromosomes (Chr) 16, 17, 19 and 20 exhibit higher PQS frequencies than those expected for their respective lengths. (**b**) Chromosomes enriched in PQS displayed increased incidence of γH2AX domains; red histograms represent the number of γH2AX domains for each chromosome observed by ChIP-Seq analyses (**Supplementary Fig. 7**); grey and black histograms represent the experimental number of γH2AX domains normalized against the theoretical number of γH2AX domains based on PQS frequencies and chromosome size, respectively. Note that Chr 16, 17, 19 and 20 contained more γH2AX domains than expected for their sizes and that this correlated with their high PQS frequencies.

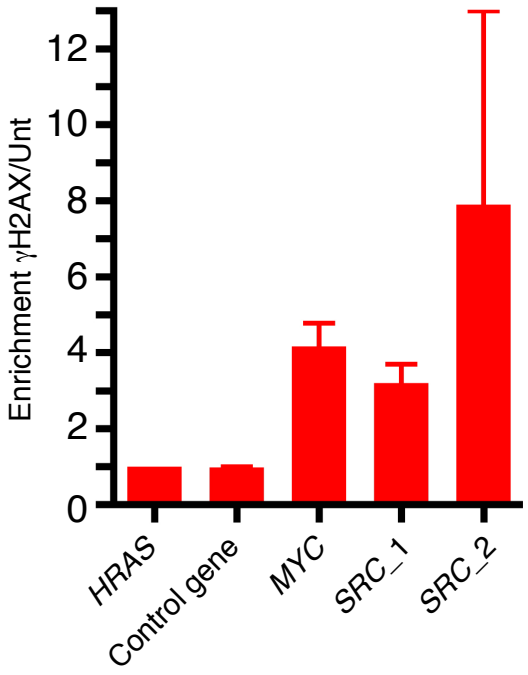


Supplementary Figure 9 | Examples of γH2AX-enriched genes following pyridostatin treatment. Data are expressed as in **Fig. 5a**; all the positive genes are found in **Supplementary Fig. 10**. At least three γH2AX-positive genes are shown for both oncogenes and tumor suppressors; *BRCA2* and *ATM* are included as examples of large control genes devoid of both PQS and γH2AX-enrichment.

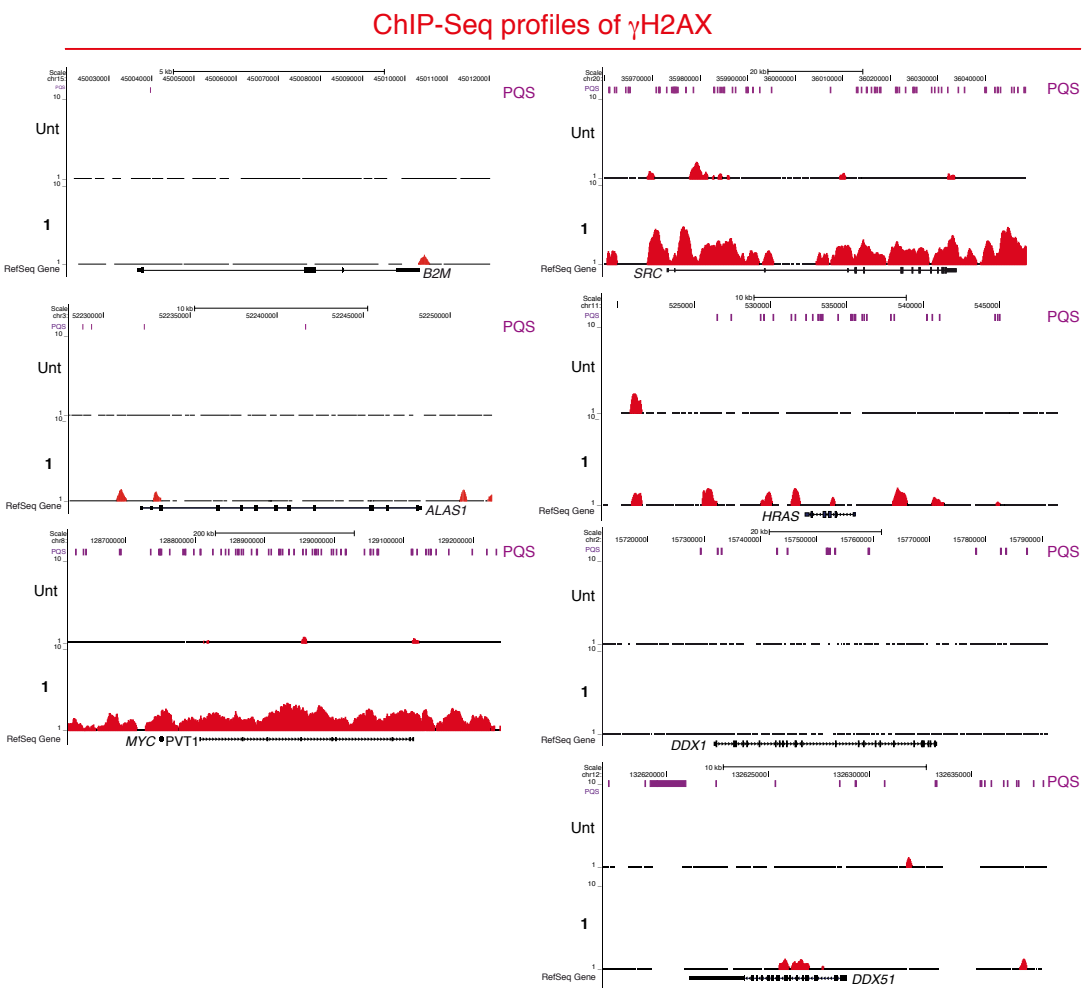
Oncogenes	% PQS in gene (Fold increase over median)
RHOC	4.96 (19.27)
SRC	3.08 (11.97)
SREBF1	2.56 (9.96)
FAIM2	2.07 (8.04)
VAV2	2.05 (7.97)
CRTC1	1.88 (7.32)
RET	1.84 (7.16)
MYC	1.74 (6.77)
WNT3	1.56 (6.08)
MAP2K3	1.49 (5.81)
WNT7A	1.49 (5.81)
SPI1	1.47 (5.72)
GLI2	0.99 (3.86)
TOM1L2	0.63 (2.44)
RAB31	0.45 (1.75)

Tumor suppressors	% PQS in gene (Fold increase over median)
CST6	5.90 (22.94)
NOTCH1	4.09 (15.89)
C9orf140	3.17 (12.33)
LLGL1	2.54 (9.87)
AHRR	2.12 (8.26)
RASSF4	1.54 (5.97)
KAT5	1.52 (5.91)
SEPT9	1.36 (5.28)
CDH4	0.99 (3.85)
DKK3	0.68 (2.63)

Supplementary Fig. 10 | Analysis of γH2AX-positive oncogenes and tumor suppressors resulting from pyridostatin treatment. Table representing top 25 genes scoring positive for γH2AX upon treatment with **1**. The % PQS for each individual gene and the fold-increase over median value for the human transcriptome are indicated.



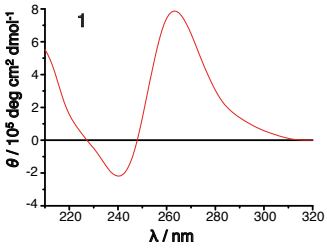
Supplementary Figure 11 | Validation of pyridostatin-induced γH2AX on target genes by ChIP-qPCR. Cells were treated with **1** and processed by ChIP as in **Fig. 5**. Following ChIP of γH2AX, samples were analyzed by quantitative PCR (qPCR) with the primers described in **Supplementary Methods**. Control gene represents the *KRT80* gene, which was devoid of γH2AX signal in the ChIP-Seq data. qPCR values were normalized to input and the control gene; data represent 3 independent experiments and error bars represent S.E.M.

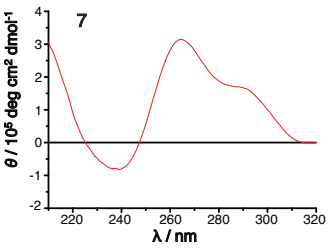


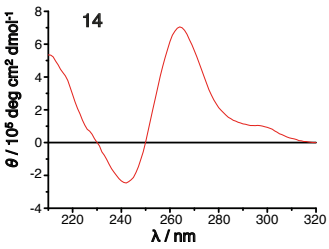
Supplementary Figure 12 | Examples of γH2AX ChIP-Seq profiles for genes analyzed by qRT-PCR. Representative profiles from the UCSC browser comparing the untreated and pyridostatin treated γH2AX ChIP-Seq profiles for several genes analyzed in **Fig. 5d**. Purple bars represent mapped PQS. The percentage of PQS bases for each gene (% PQS in gene) of the transcriptome is given in **Supplementary Dataset 3**.

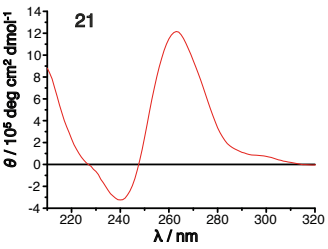
Sequence Number	Strand	Position in the Gene	Sequences (5´ to 3´)	G-quadruplex
1	+	373-404	–GGGTGACTTGGGTGTCCGGGGGGTGGGGGGG–	+
2	+	628-657	–GGGGCAGCTGGGTCGCTCGGGGAACGGGG–	+
3	+	1348-1366	–GGGAGGGAGGGCTGGGGG–	+
4	+	2341-2361	–GGGCGGCGGGCTGGGCGGGG–	+
5	+	3398-3417	–GGGGGGGTGGGAGGGGGGG–	+
6	+	4128-4146	–GGGAGGGTGGGCTGGGGG–	+
7	+	4228-4253	–GGGAGATGGGGGTGGGTGGATTGGG–	+
8	+	4653-4677	–GGGAAAGGGATTTGGGGGGCGGGG–	+
9	+	10652-10678	–GGGCCTGGGGCTCAGCGGGAGATGGG–	+
10	+	12174-12192	–GGGCGGAGGGCGGGCGGG–	+
11	+	12619-12640	–GGGTGAGGGCTCGGGCGTGGG–	n.d.
12	+	16150-16175	–GGGCGGCCGGGGCAGGGGGCAGGGG–	+
13	+	18175-18194	–GGGGAGGGTGGGGAAGGGG–	+
14	–	2573-2602	–GGGTGGGCTGACGGGGCAGAATGGGAGGG–	+
15	–	5132-5158	–GGGTGGGAGGAAGGGGCAGTTCAGGG–	+
16	–	5451-5486	–GGGCTGCTGGGACTGGTTGGGAGGGCACTCTGGGG–	+
17	–	8570-8595	–GGGCAGGGAGACCTGGGAACTAGGG–	+
18	–	8612-8637	–GGGGACAGGGACGGGGTCCACAGGG–	+
19	–	8871-8905	–GGGTCCCCTGGGCCTGGGCGGGAGCGGGGAGGGG–	+
20	–	9368-9383	–GGGAGGGAGGGAGGG–	+
21	–	12268-12292	–GGGAGGGGGGAGGGGGAAGTGGGG–	+
22	–	12348-12384	–GGGAGAGAGGGGGACAGGGAGGAGGGGGGCTGGGGG–	n.d.
23	–	12982-13009	–GGGGCGCAGGGAGCAGGTGGGGAAGGG–	+
24	–	16715-16741	–GGGTGCTTGGGTTTGGGGCTGGAGGG–	+
25	–	17534-17556	–GGGTGTGGGGAGGCGGGAGGGG–	+

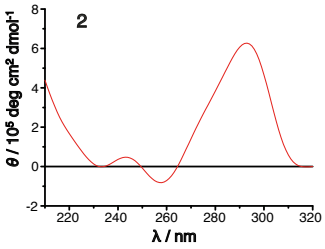
Supplementary Figure 13 | List of PQS identified in the coding region of *SRC*. The symbol + of the G-quadruplex-labelled column indicates that both CD and NMR spectroscopic methods confirmed the formation of a G-quadruplex structure *in vitro*; n.d. (not determined) indicates that either CD or NMR spectroscopy was not conclusive.

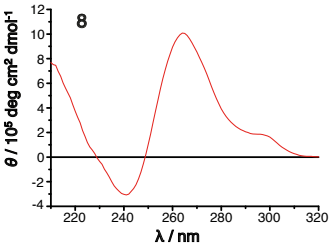

1
θ / 10⁵ deg cm² dmol⁻¹
λ / nm

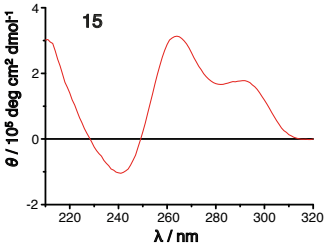

7
θ / 10⁵ deg cm² dmol⁻¹
λ / nm

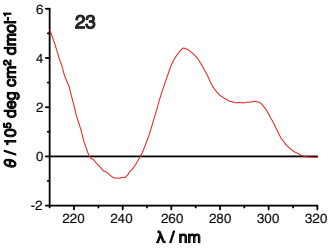

14
θ / 10⁵ deg cm² dmol⁻¹
λ / nm

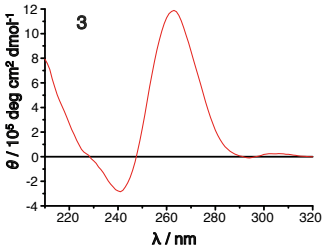

21
θ / 10⁵ deg cm² dmol⁻¹
λ / nm

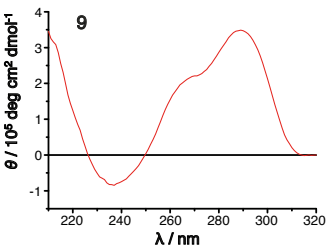

2
θ / 10⁵ deg cm² dmol⁻¹
λ / nm

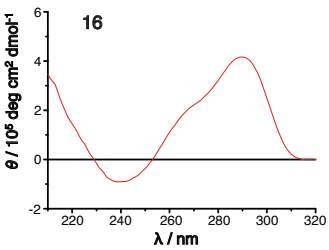

8
θ / 10⁵ deg cm² dmol⁻¹
λ / nm

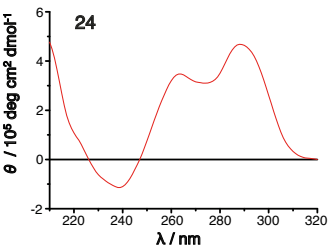

15
θ / 10⁵ deg cm² dmol⁻¹
λ / nm

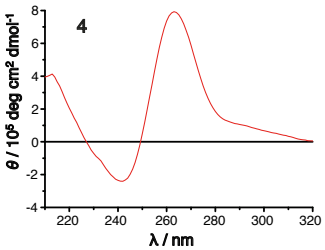

23
θ / 10⁵ deg cm² dmol⁻¹
λ / nm

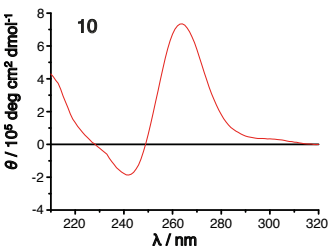

3
θ / 10⁵ deg cm² dmol⁻¹
λ / nm

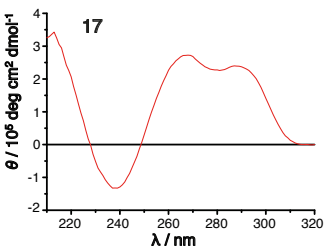

9
θ / 10⁵ deg cm² dmol⁻¹
λ / nm

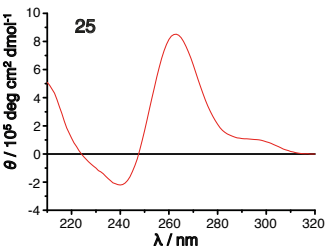

16
θ / 10⁵ deg cm² dmol⁻¹
λ / nm

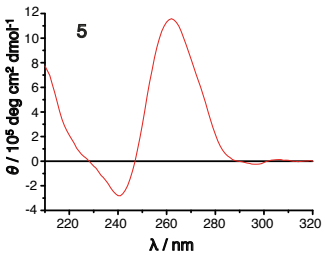

24
θ / 10⁵ deg cm² dmol⁻¹
λ / nm

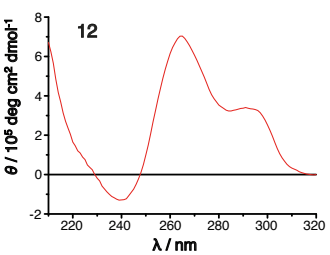

4
θ / 10⁵ deg cm² dmol⁻¹
λ / nm

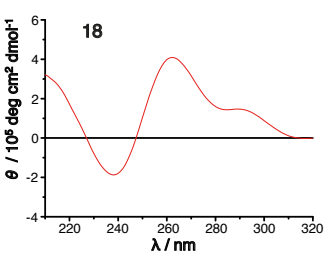

10
θ / 10⁵ deg cm² dmol⁻¹
λ / nm

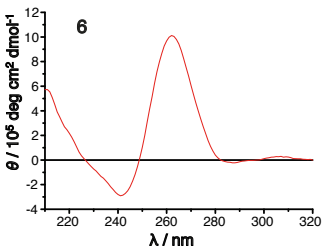

17
θ / 10⁵ deg cm² dmol⁻¹
λ / nm

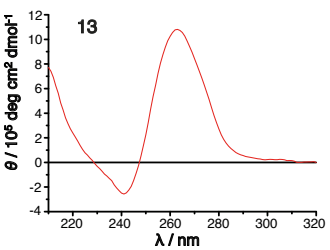

25
θ / 10⁵ deg cm² dmol⁻¹
λ / nm

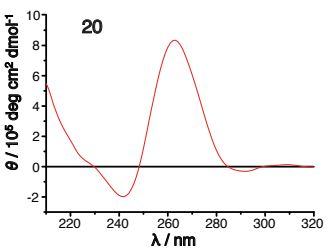

5
θ / 10⁵ deg cm² dmol⁻¹
λ / nm


12
θ / 10⁵ deg cm² dmol⁻¹
λ / nm

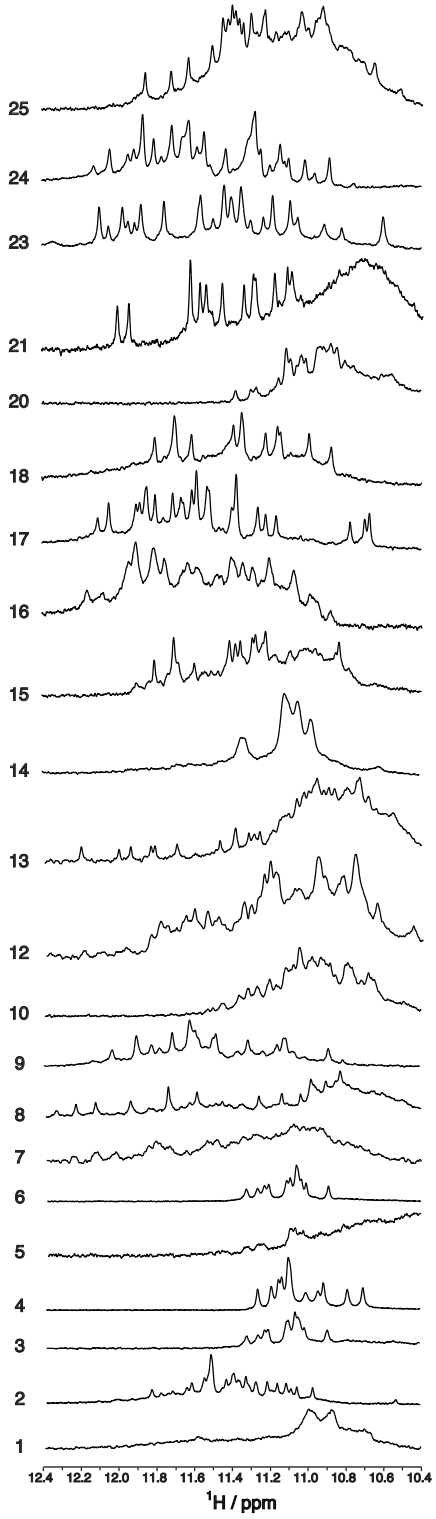

18
θ / 10⁵ deg cm² dmol⁻¹
λ / nm


6
θ / 10⁵ deg cm² dmol⁻¹
λ / nm

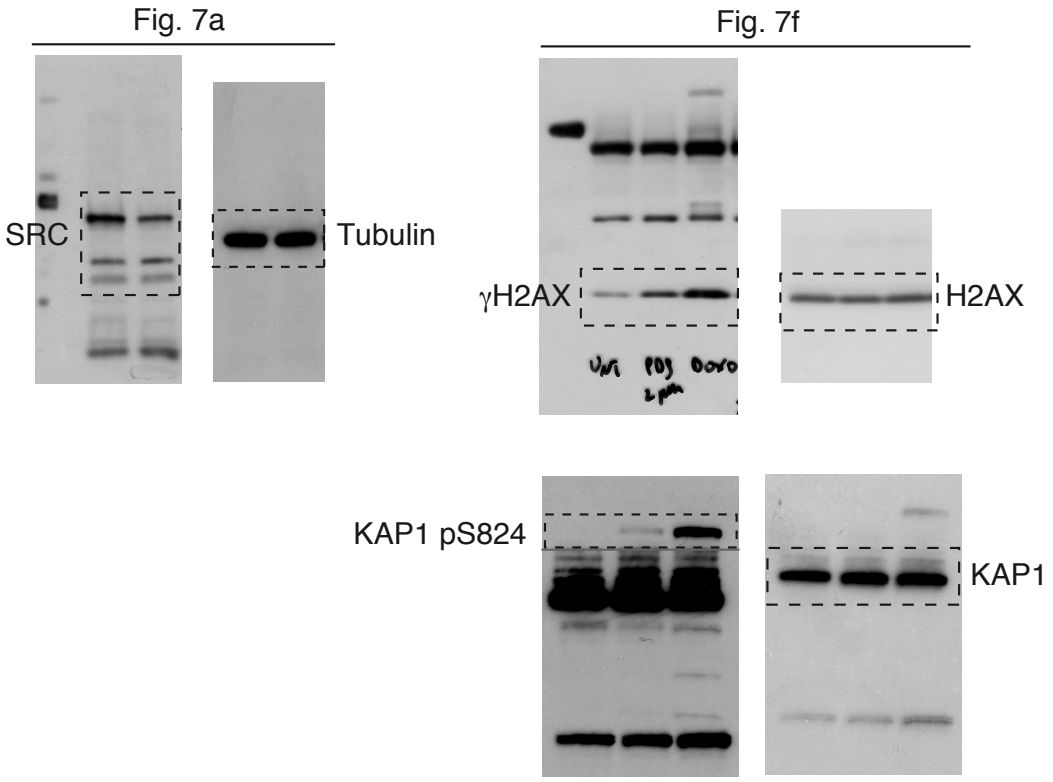

13
θ / 10⁵ deg cm² dmol⁻¹
λ / nm


20
θ / 10⁵ deg cm² dmol⁻¹
λ / nm

Supplementary Figure 14 | CD spectra of PQS identified in the coding region of *SRC*. The molar ellipticity of each individual PQS pre-annealed in potassium containing buffer displays signals at 265 nm and/or 298 nm characteristic of parallel G-quadruplex conformations and/or antiparallel conformations, respectively. Several spectra exhibit both signals, demonstrating the presence of a mixture of conformers or the presence of mixed-type parallel/antiparallel conformations.



Supplementary Figure 15 | NMR spectra of PQS identified in the coding region of *SRC*. The chemical shift of each individual PQS pre-annealed in potassium containing buffer exhibits characteristic imino proton signals shifted downfield between 10.5 to 12.5 ppm characteristic of Hoogsteen hydrogen bonding and G-quadruplex structure formation.



Supplementary Figure 16 | Full gel images for blots shown in Fig. 7a,f. After electrophoresis and transfer, blots were cut into strips that contained the protein of interest and were stained with the indicated antibodies. Dashed boxes represent the regions displayed in **Fig 7a,f**. Note that SRC and Tubulin were individually analyzed as single blots.

SUPPLEMENTARY DATASET LEGENDS

These datasets can be accessed by the following weblinks:

http://bioinformatics.gurdon.cam.ac.uk/Rodriguez/Rodriguez_Supplementarydataset_1.xlsx

http://bioinformatics.gurdon.cam.ac.uk/Rodriguez/Rodriguez_Supplementarydataset_2.xls

http://bioinformatics.gurdon.cam.ac.uk/Rodriguez/Rodriguez_Supplementarydataset_3.xls

Supplementary Dataset 1 | Mapped PQS across the human genome. Each individual PQS is identified by chromosome number (column A) with its starting (column B) and ending (column C) genomic position indicated, and whether the PQS is found on the plus (+) or minus (+) strand (column D).

Supplementary Dataset 2 | γH2AX-enrichment score of oncogenes and tumor suppressors (as described by CancerGenes). Tables are divided into three sections containing the Entrez ID number (column A), Gene Symbol (column B), % PQS in gene (column C; as determined in **Supplementary Dataset 3**) and γH2AX score (column D); protocol for scoring γH2AX-enrichment is described in **Methods**. **Supplementary Dataset 2** is divided into three sheets with the summary, oncogenes and tumor suppressors respectively.

Supplementary Dataset 3 | Size, number, strand location and distribution of all PQS in individual genes of the transcriptome. For each gene, the RefSeq ID (column A), gene symbol (column B), its chromosome location (column C) and transcribed strand (column D) are provided, the size of each gene (column G) and its exon count (column I) are also given. For PQS, the number of motifs (column E), size (column F), strand location (column J for sense and column K for antisense) and sense/antisense ratio have been calculated and are shown.

SUPPLEMENTARY REFERENCE

51. Molander, G.A. & Cormier, E.P. Ketyl-allene cyclizations promoted by samarium(II) iodide. *J. Org. Chem.* **70**, 2622-6 (2005).

52. Lees-Miller, S.P. *et al*. Absence of p350 subunit of DNA-activated protein kinase from a radiosensitive human cell line. *Science* **267**, 1183-5 (1995).

53. Cogoi, S. & Xodo, L.E. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.* **34**, 2536-49 (2006).