

SFPQ directs histone H3.3 deposition to R-loops in DNA repeats to protect genome stability

Received: 27 December 2024

Accepted: 2 February 2026

Published online: 24 February 2026

 Check for updates

Alessandro Ferrando ^{1,5,8}, Michele Giaquinto^{1,8}, Luisa M. R. Napolitano², Giulia Canarutto ^{1,3}, Alessandro Framarini ¹, Alice Gambelli¹, Pamela Veneziano Broccia ^{1,6}, Annie Zappone¹, Eleonora Petti ^{1,7}, Chiara Boncristiani¹, Andrea Parlante^{1,3}, Silvia Onesti ², Silvano Piazza ^{1,3}, Roberta Benetti ⁴  & Stefan Schoeftner ¹ 

R-loops are three-stranded nucleic acid structures composed of an RNA:DNA hybrid duplex and a displaced single-stranded DNA loop. Unscheduled or persistent R-loops drive genome instability by creating conflicts with transcription and replication. Up to 75% of the human genome comprises repetitive DNA elements that are prone to R-loop formation. We show that the RNA binding protein SFPQ suppresses R-loop mediated replication stress and DNA damage at repeat elements such as telomeres, (peri)-centromeres, LINE-1 and SINE elements. SFPQ exhibits in-vitro R-loop binding activity, associates with chromatin containing R-loops, and recruits the histone H3.3 specific chaperon DAXX to preserve a correct nucleosome template that counteracts R-loop accumulation. Loss of SFPQ results in DAXX displacement from repeat elements, reduced histone H3.3 incorporation, replication stress-mediated genome instability and the emergence of cytoplasmic DNA. This leads to activation of innate immune signaling via the cGAS/STING pathway, ultimately correlating with improved survival of sarcoma patients.

R-loops are atypical three-stranded nucleic acid structures comprising a 60–2000 nucleotide RNA:DNA hybrid and a displaced single-stranded DNA loop^{1,2}. R-loops are estimated to occupy up to 5% of the human genome and regulate diverse physiological processes, including chromatin organization, DNA methylation, transcriptional termination, immunoglobulin class switching, sister chromatid segregation, and mitochondrial DNA replication^{3–5}. In contrast, unprogrammed or persistent R-loops pose a major threat to genome stability by generating transcription–replication conflicts, impairing DNA repair, and serving as substrates for mutagenesis^{4,6}. R-loop formation

and resolution are orchestrated by multiple machineries acting at the RNA, DNA, and chromatin level. The THO complex, SRSF1, and related factors suppress hybridization of nascent transcripts to the DNA template^{7,8}. Ribonucleases RNaseH1 and RNaseH2 degrade RNA components of R-loops⁹. ATP-dependent helicases such as DHX9, the Bloom's syndrome helicase (BLM), Aquarius (AQR), FANCM and members of the DEXD/H family of RNA helicases or chromatin regulatory factors such as the FACT and SWI/SNF chromatin remodeling complex, INO80 and DAXX/ATRAX histone chaperons prevent unprogrammed R-loops^{4,10–14}. Finally, factors that control replication fork

¹Dipartimento di Scienze della Vita, Università degli Studi di Trieste, Trieste, Italy. ²Structural Biology Laboratory, Elettra—Sincrotrone Trieste S.C.p.A, Area Science Park Basovizza, Trieste, Italy. ³Computational Biology, International Centre for Genetic Engineering and Biotechnology, Trieste, Italy. ⁴Laboratory of Epigenomics, Department of Medicine, Università degli Studi di Udine, Udine, Italy. ⁵Present address: Dipartimento di Scienze Cliniche e Biologiche, Università degli Studi di Torino, Orbassano, Italy. ⁶Present address: Centro Nacional de Análisis Genómico (CNAG), Barcelona Science Park—Tower I, Barcelona, Spain. ⁷Present address: Translational Oncology Research Unit, IRCCS—Regina Elena National Cancer Institute, Rome, Italy. ⁸These authors contributed equally: Alessandro Ferrando, Michele Giaquinto. ✉e-mail: roberta.benetti@uniud.it; sschoeftner@units.it

stability—including BRCA1/2, Fanconi anemia proteins, topoisomerases or DNA damage signaling factors such as Ataxia telangiectasia and Rad3 related (ATR) prevent R-loop-mediated genome instability^{15–22}.

Privileged sites for R-loop-mediated genome instability are characterized by GC- and AT-skew, G-richness or arrangement in tandem repeats^{23–26}. Accordingly, R-loops were found to be enriched at telomeric and (peri-)centromeric satellite repeats, simple/low-complexity repeats, rDNA repeat arrays and retroelements^{25,27–34}. Mutations in canonical R-loop suppression factors induce instability of DNA repeats, a central feature of several genetic disorders²⁵. Immunodeficiency, centromeric instability, and facial anomalies syndrome type I is marked by elevated R-loop levels and instability at (peri)centric, telomeric, and rDNA repeats³⁵. Patients with Bloom syndrome or Ataxia-telangiectasia mutated deficiency exhibit instability of rDNA arrays³⁶. Mutations or dysregulated expression of the *ATRX* or *DAXX* genes compromise retroelement stability, driving R-loop-mediated replication stress and increased recombination at telomere repeats^{13,29,37–39}. The pervasive transcriptional activity across the human genome necessitates R-loop surveillance machineries that safeguard genome integrity within repetitive elements, which account for ~75% of the genome^{40,41}. However, surveillance mechanisms that detect and resolve unprogrammed R-loops at repetitive elements remain poorly defined. SFPQ is a multifunctional RNA-binding protein that, together with NONO and PSP1, belongs to the conserved DBHS (Drosophila behavior/human splicing) protein family⁴². DBHS proteins form homo- and heterodimers through their NOPS and RRM2 domains and are enriched in paraspeckles in vertebrates. Beyond paraspeckle assembly, SFPQ has been implicated in transcriptional regulation, splicing, DNA damage repair, stress responses, and telomere maintenance^{43–51}.

Here we demonstrate a critical role for the RNA-binding protein SFPQ in preventing aberrant R-loop formation across a broad spectrum of repetitive elements, including pericentromeric and centromeric satellite repeats, subtelomeres, telomeres, as well as LINE-1 and SINE elements. SFPQ displays binding specificity for R-loop structures in vitro, associates with chromosomal R-loops via its RRM domains, and employs its proline-rich domain to recruit the histone H3.3-specific chaperone DAXX. This recruitment preserves nucleosome architecture at repetitive elements, thereby counteracting R-loop-mediated genome instability and limiting the accumulation of cytoplasmic DNA. Disruption of this pathway induces genome instability in cancer cells and activates innate immune signaling, which correlates with improved survival in patients with sarcoma.

Results

SFPQ targets R-loops in repetitive elements

SFPQ suppresses R-loop formation at vertebrate telomeric repeats, thereby protecting chromosome ends from replication stress, DNA damage, and excessive recombination events⁴⁶. Immuno-dot blot analysis of genomic DNA from SFPQ-depleted U-2 OS, H1299, MCF-7, and MCF-10A cells revealed a significant increase in steady-state DNA:RNA hybrid levels, consistent with a role for SFPQ in R-loop suppression at multiple loci (Supplementary Fig. 1a, b). To extend this link to repetitive elements, we performed RNA-FISH under non-denaturing conditions using probes derived from human C₀t-1 DNA, which is enriched for interspersed repetitive sequences⁵². Both the number and signal intensity of C₀t-1 RNA-FISH foci increased significantly following depletion of SFPQ or RNaseH1 in U-2 OS and H1299 cells (Fig. 1A, Supplementary Fig. 1c). This finding parallels the reported increase in focal TERRA staining intensity under conditions of elevated telomeric R-loop abundance induced by loss of SFPQ or RNaseH1^{29,46}. Specificity of used short interfering RNAs for *SFPQ* and *RNaseH1* mRNA was validated by western blotting (Supplementary Fig. 1d, Supplementary Fig. 7j, see below). Importantly, C₀t-1 RNA-FISH signals in SFPQ- or RNaseH1-depleted U-2 OS cells co-localized with p-ATR

(Thr1989) and FANCD2, indicative of replication stress at repetitive elements (Fig. 1B, Supplementary Fig. 1e). Although C₀t-1 RNA-FISH lacks sufficient sensitivity to comprehensively monitor transcriptional activity across all repeat sequences, our findings demonstrate that loss of RNaseH1 or SFPQ induces aberrant RNA metabolism at repeat-containing loci, thereby promoting replication stress. Consistent with ATR activation, SFPQ depletion resulted in elevated phosphorylation of Chk1 and RPA32, as shown by western blot analysis (Supplementary Fig. 1f). The induction of replication stress upon SFPQ depletion is consistent with defects in S-phase and G₂-M progression and previous reports (Supplementary Fig. 1g)^{53–55}. SFPQ was found to accumulate at centromeric and telomeric repeats following experimental elevation of R-loop levels induced by transient RNaseH1 depletion or hydroxyurea treatment, as shown by co-immunofluorescence with CREST/SFPQ and TRF2/SFPQ antibodies (Fig. 1C, D; Supplementary Fig. 1h–j). Chromatin immunoprecipitation in RNaseH1-depleted U-2 OS cells further demonstrated recruitment of SFPQ to diverse classes of repetitive elements, including telomeric repeats, subtelomeres of Chr16p, Sat III, Sat II, centromeric α -satellite repeats, AluS, AluY, and LINE elements (Fig. 1E). Thus, elevated R-loop levels at major categories of vertebrate DNA repeats drive dynamic SFPQ recruitment. To assess whether SFPQ contributes to R-loop suppression at these loci, we transiently depleted SFPQ in U-2 OS cells and performed DNA:RNA immunoprecipitation (DRIP). Loss of SFPQ resulted in increased R-loop abundance across telomeric repeats, subtelomeres of Chr16p, Sat III, Sat II, centromeric α -satellite repeats, AluS, AluY and LINE elements, as well as at a known SFPQ binding site in the promoter of the *PD3A* gene⁵⁶ (Fig. 1F). Negative control regions in the *WRNIP* and *SSH* promoters did not show R-loop formation. Consistent with these findings, RNaseH1 depletion also led to elevated R-loop levels at all classes of repetitive elements and at the *PD3A* promoter (Supplementary Fig. 1k). Together, these data indicate that SFPQ is recruited to unscheduled R-loops at repetitive elements and exerts R-loop-suppressive activity.

SFPQ has in vitro binding specificity for R-loops

We next examined the binding activity of SFPQ towards distinct, 5'-[6-FAM] labeled nucleic acid substrates, including single-, double- and triple-stranded structures, such as RNA:DNA hybrids, D-loops and R-loops (Supplementary data 1). Full-length human SFPQ carrying an N-terminal His₆-myc tag was purified as recombinant protein from *E. coli* and tested in electrophoretic mobility shift assays (EMSA). Native polyacrylamide electrophoresis revealed that recombinant SFPQ does not interact with dsDNA or dsRNA, but binds ssDNA and shows stronger affinity for ssRNA (Fig. 2A, B; Supplementary Fig. 2a). Binding to single-stranded, non-repetitive RNA was enhanced with increasing substrate length from 15 to 23 nucleotides, indicating a preference for longer RNA molecules (Supplementary Fig. 2b). Interestingly, SFPQ displayed markedly increased affinity for repeat-containing single-stranded RNAs, including [UUAGGG]₄ and [GGAAU]₄ repeats present in TERRA and pericentric human SatIII RNAs, respectively, as well as [GAAA]₅ repeats (Supplementary Fig. 2b). These findings indicate that SFPQ preferentially binds RNA repeats in vitro. Consistent with this observation, SFPQ was detected in eluates from TERRA RNA pull-down assays using nuclear protein extracts⁴⁶. We next assessed the binding capacity of SFPQ towards atypical nucleic acid structures. Recombinant SFPQ did not interact with a fully paired DNA:RNA hybrid duplex or with dsDNA, but bound a non-repetitive dsDNA probe containing a central unpaired region (DNA bubble; Fig. 2B). Binding was more efficient with a synthetic R-loop probe lacking a protruding RNA tail (Fig. 2B), and was further enhanced with D-loop structures carrying 5' or 3' single-stranded DNA termini (Fig. 2C). Most efficient mobility shifts were observed using R-loop substrates containing single stranded 5' or 3' termini RNA termini that protrude from the loop structure (R-loop 3'tails, R-loop 5'tails; Fig. 2D, E;

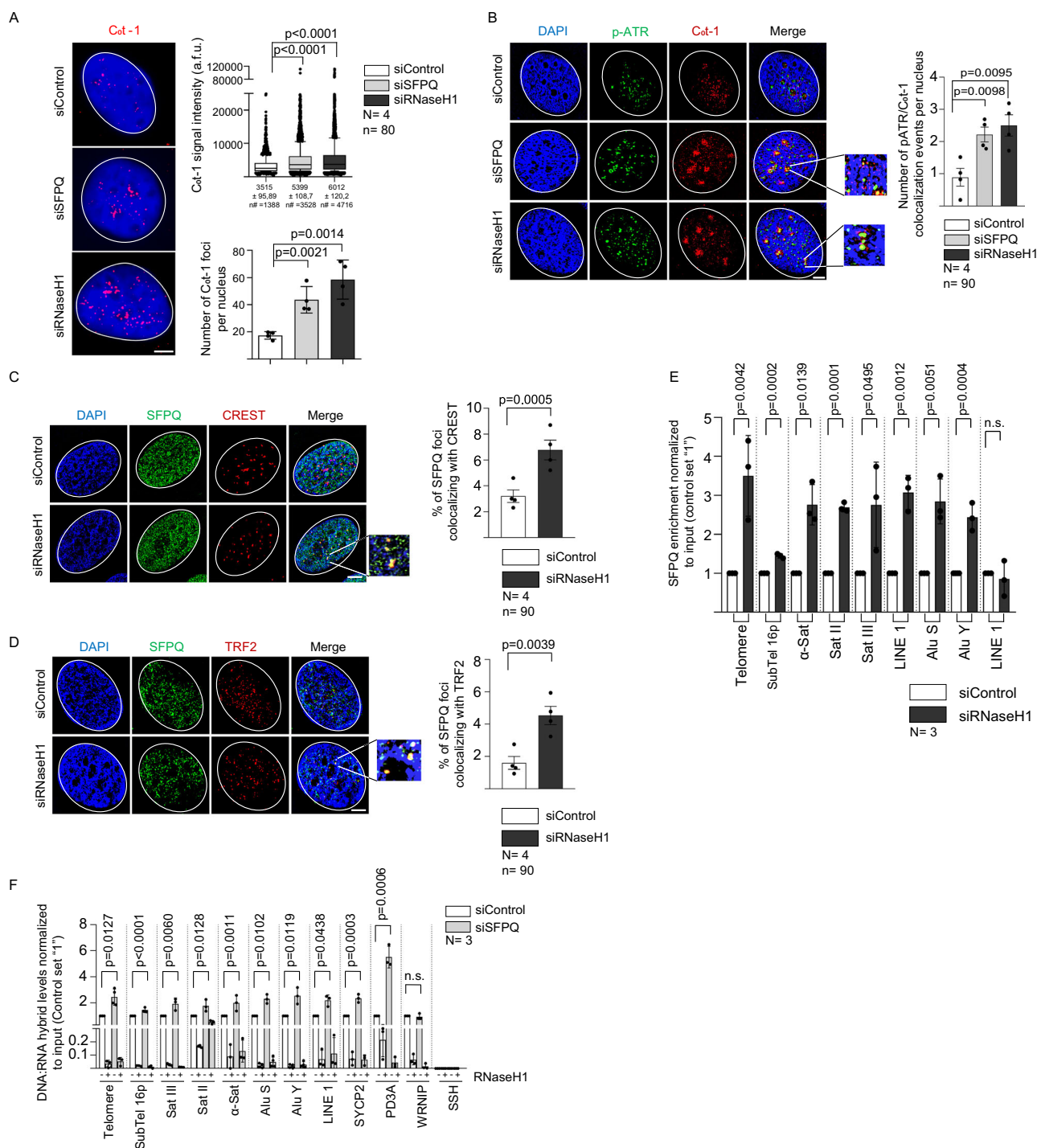


Fig. 1 | SFPQ suppresses R-loop formation in repeat elements. **A** COT-1 RNA-FISH performed in U-2 OS cells transiently transfected with indicated siRNAs (left panel). Quantification of focal COT-1 signal intensity (right, top panel) and number of COT-1 foci per nucleus (right bottom panel). **B** COT-1 RNA-FISH combined with anti-p-ATR (Thr1989) immunostaining performed in U-2 OS cells transiently transfected with indicated siRNAs. Images were obtained by super-resolution microscopy. Right panel, quantification of number of COT-1 foci co-localizing with p-ATR (Thr1989) per nucleus. **C** Combined immunofluorescence with anti-SFPQ and anti-CREST (centromere) antibodies in U-2 OS cells transfected with indicated siRNAs. Images were obtained by super-resolution microscopy. Right panel, quantification of co-localization events. **D** Representative images of combined immunofluorescence with anti-SFPQ and anti-TRF2 antibodies in U-2 OS cells transfected with indicated siRNAs. Images were obtained by super-resolution microscopy. Right panel,

quantification of co-localization events. **E** Anti-SFPQ ChIP-qPCR analysis using control and RNaseH1 depleted U-2 OS cells. Indicated sequence categories were amplified using specific PCR primer pairs. **F** Quantitative DRIP-PCR analysis after siRNA mediated SFPQ knock-down in U-2 OS cells. RNaseH1, treatment of genomic DNA with recombinant RNaseH1 prior to DRIP. Indicated repeat regions were amplified using specific primers. Mean values are shown, error bars indicate standard deviation. *N* = number of independent experiments. *n* = number of analyzed nuclei. An unpaired, two-sided Student's *t* test was used to calculate statistical significance; *p*-values are shown. Scale bar (1 μm) applies to all images in respective immunofluorescence panels. DNA was stained using DAPI (4',6-diamidino-2-phenylindole). Arrowheads and zoomed images indicate co-localization events. Source data are provided with this paper.

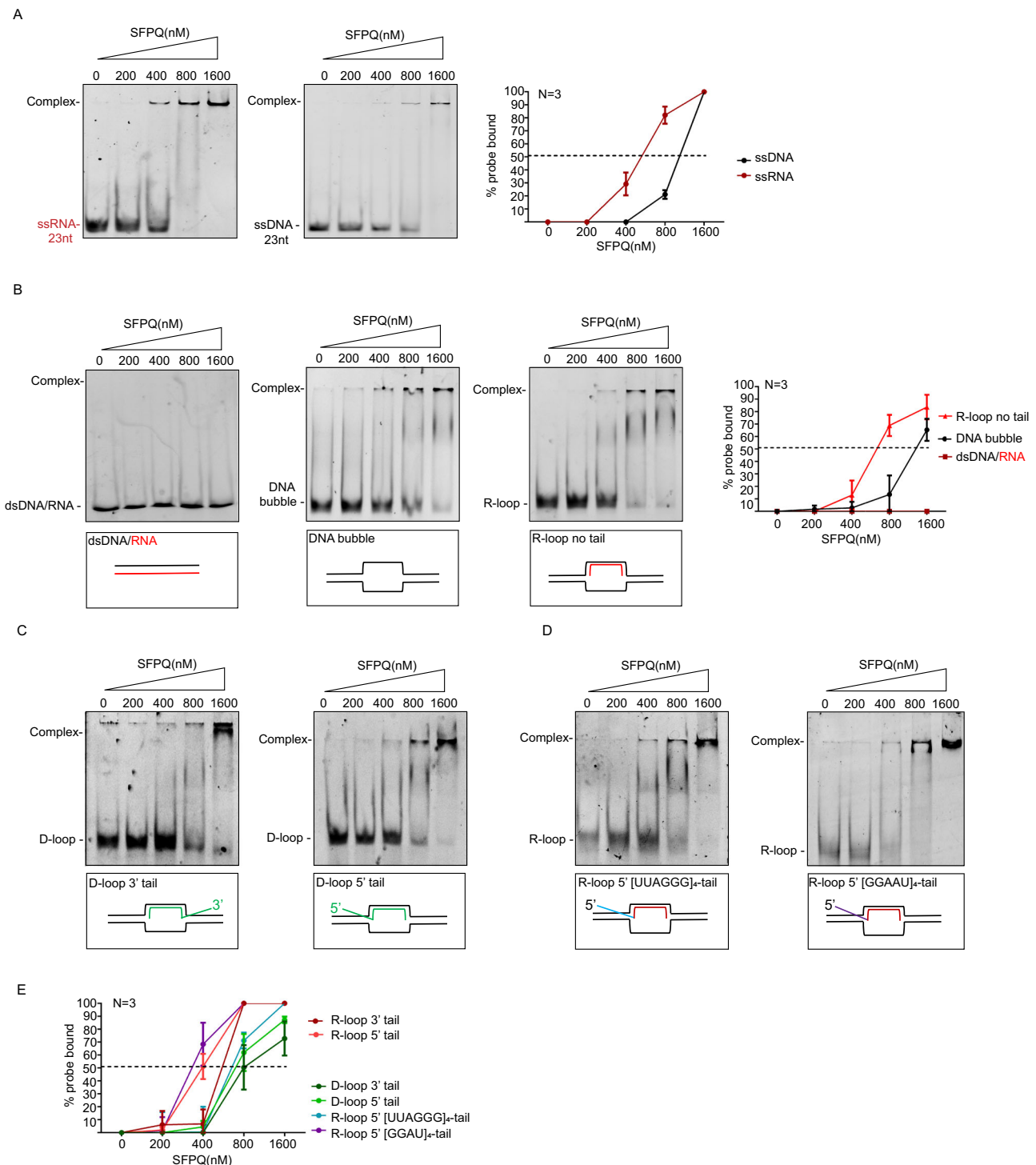


Fig. 2 | SFPQ binds R-loop structures in vitro. **A** EMSA showing recombinant SFPQ binding to ssDNA (left) and ssRNA (middle). Right panel, quantification of shifted ssRNA and ssDNA substrates. **B** Representative images of EMSA showing recombinant SFPQ binding to dsDNA:RNA duplexes (left), DNA bubble (center-left) and R-loop without protruding termini (center-right). Right panel, quantification of shifted nucleic acid substrates. **C** Representative images of EMSA showing recombinant SFPQ binding to D-loop substrate with protruding 3' ssDNA terminus

(left) or 5' ssDNA terminus (right). **D** Representative images of EMSA showing recombinant SFPQ binding to R-loop substrate with protruding 5' ends composed of single stranded RNA repeats ([UUAGGG]₄ or [GGAAU]₂). **E** Quantification of EMSA assays shown in **C**, **D** and Supplementary Fig. 2c. **A**, **B**, **E** Data are represented as mean values; error bars indicate standard deviation; *N* = number of independent experiments. Source data are provided with this paper.

Supplementary Fig. 2c). To evaluate the contribution of these ssRNA protrusions to SFPQ binding, we performed competition assays with unlabeled ssRNA or dsDNA. As expected, dsDNA did not interfere with R-loop binding. Remarkably, tenfold molar excess of ssRNA was

required to disrupt SFPQ association with R-loops containing ssRNA protrusions (Supplementary Fig. 2d). Together, these findings demonstrate that SFPQ exhibits in vitro specificity for three-stranded R-loop structures.

SFPQ is a binding partner of the Death Domain Associated Protein (DAXX)

To date, no enzymatic activity has been attributed to SFPQ. To identify SFPQ-interacting proteins that may influence R-loop metabolism, we transiently overexpressed myc-tagged human SFPQ in the highly transfectable HI299 lung adenocarcinoma cell line and performed anti-myc immunoprecipitation. Control immunoprecipitations were carried out using myc-tagged Bax-interacting factor 1 (Bif-1)⁵⁷. Coomassie staining revealed distinct bands in myc-SFPQ eluates, which were excised from the protein gel and analyzed by mass spectrometry, identifying a panel of candidate SFPQ-interacting proteins (Fig. 3A; Supplementary data 2–6). The detection of NONO, a previously reported SFPQ interactor, validated the specificity of our experimental approach (Fig. 3A; Supplementary data 6). Given its established role in genome stability, we directed subsequent analyses toward the Death Domain–Associated Protein (DAXX) (Supplementary data 2). DAXX is a multifunctional protein involved in apoptosis, senescence, gene transcription, and the DNA damage response⁵⁸. In addition, DAXX functions as a histone chaperone, ensuring the incorporation of the histone variant H3.3 into repeat regions to preserve genome stability^{37,59}. The SFPQ–DAXX interaction was validated by reciprocal co-immunoprecipitation of ectopically expressed myc-tagged SFPQ and HA-tagged DAXX in HI299 cells (Fig. 3B, C). Immunoprecipitation of endogenous proteins further demonstrated that a relevant fraction of total SFPQ and DAXX resides within a protein complex in U-2 OS cells (Fig. 3D). Direct interaction between SFPQ and DAXX was confirmed by pull-down assays using recombinant His₆–myc-tagged SFPQ and GST-tagged DAXX–HA proteins (Supplementary Fig. 3a, b). Western blotting further detected DAXX in reverse-crosslinked chromatin isolated from anti-SFPQ ChIP experiments (Supplementary Fig. 3c). Consistent with SFPQ knockdown data, siRNA-mediated depletion of DAXX in U-2 OS cells increased global R-loop levels and ATR phosphorylation (Supplementary Fig. 3d–f). This phenotype was rescued by inducible RNaseH1 expression, supporting the conclusion that both SFPQ and DAXX are critical for R-loop regulation in our model system (Supplementary Fig. 3D–F). The specificity of the S9.6 antibody was validated by pretreating fixed cells with RNaseT1, RNaseT3, or RNaseH1 (Supplementary Fig. 3e). Consistent with the observation that SFPQ is recruited to ectopic R-loops (Fig. 1C–F; Supplementary Fig. 1i–k), RNaseH1 knockdown increased DAXX localization at telomere repeats (Supplementary Fig. 3g). Additional immunofluorescence analyses revealed that SFPQ depletion in U-2 OS and HI299 cells produced a dispersed nuclear DAXX distribution, accompanied by reduced DAXX localization at telomeric and centromeric regions, both known R-loop hotspots (Fig. 3E, F; Supplementary Fig. 3m). Loss of SFPQ led to an ~20% increase in DAXX protein and mRNA levels in U-2 OS and HI299 cells (Supplementary Fig. 3h–k). DAXX is known to interact with ATRX to form a complex that enforces repressive chromatin architecture at repeat elements³⁷. ATRX is epigenetically silenced in U-2 OS cells but expressed in HI299 cells (Supplementary Fig. 3i)⁶⁰. Notably, SFPQ depletion in HI299 cells caused a pronounced reduction in steady-state ATRX protein levels, while only modestly affecting *ATRX* mRNA expression (Supplementary Fig. 3i–l). Treatment of SFPQ-depleted HI299 cells with MG132 restored ATRX protein levels (Supplementary Fig. 3n), indicating that ATRX is targeted for proteasomal degradation in the absence of SFPQ. We conclude that SFPQ regulates DAXX and ATRX function in cancer cells by directing DAXX localization and sustaining ATRX protein expression.

SFPQ controls DAXX-dependent histone H3.3 deposition

To elucidate the role of SFPQ in orchestrating DAXX-dependent deposition of histone H3.3, we performed Chromatin Immunoprecipitation followed by sequencing (ChIP–seq) to map the genomic localization of SFPQ, DAXX, and histone H3.3 in control and SFPQ-depleted U-2 OS cells. In control cells, ChIP–seq identified 6375 peaks

for SFPQ and 6,643 peaks for DAXX. Consistent with the observation that DAXX-associated chromatin is co-immunoprecipitated using SFPQ-specific antibodies (Supplementary Fig. 3c), DAXX occupancy was markedly reduced in SFPQ knockdown cells, with only 2816 peaks detected (Supplementary Fig. 4a). In control U-2 OS cells, SFPQ, DAXX, and histone H3.3 peaks were broadly distributed across all annotated sequence categories (Supplementary Fig. 4b, c). Notably, unlike SFPQ, a substantial proportion of DAXX and histone H3.3 peaks was enriched in regulatory regions, including promoters and transcription start sites (TSS; 9.54% and 5.24%, respectively), as well as CpG-rich domains (1.99% and 1.02%, respectively) (Supplementary Fig. 4c). To explore the potential role of SFPQ in genome-wide recruitment of DAXX, we focused on loci co-occupied by both factors. In control cells, 1331 of the 5168 total DAXX peaks overlapped with SFPQ binding sites (Supplementary Fig. 4d). SFPQ–DAXX co-occupied peaks were largely excluded from promoters, transcription start sites (TSS; 0.39%), and CpG-rich regions (0.26%), and were predominantly located within intergenic regions, introns, and repetitive elements including LINEs, SINEs, satellites, and simple repeats (Fig. 4A; Supplementary Fig. 4c). Consistent with immunofluorescence data, SFPQ depletion markedly altered the genome-wide distribution of DAXX (Fig. 4B). Specifically, DAXX occupancy at promoter–TSS regions increased from 9.54% in control cells to 21.12% following SFPQ knockdown, and from 1.99% to 5.34% in CpG-rich domains (Fig. 4B; Supplementary Fig. 4c). A comparable trend was found for 5'UTR and exonic sequences (Fig. 4B, Supplementary Fig. 4c). Increased DAXX occupancy at regulatory elements was paralleled by reduced frequency of DAXX peaks at intronic, LINE, SINE and intergenic sequences (Supplementary Fig. 4c, e). Although SFPQ depletion led to a marked reduction in the total number of DAXX peaks, significant log₂-fold changes in the remaining DAXX peaks were confined to promoter–TSS regions, CpG-rich sequences, intergenic domains, and LINE elements (Supplementary Fig. 4e). These findings underscore the requirement of SFPQ for proper genome-wide distribution of DAXX. To assess whether aberrant DAXX localization in SFPQ-deficient cells affects histone variant deposition, we examined H3.3 occupancy. Loss of SFPQ resulted in a pronounced increase in histone H3.3 peak numbers, rising from 29,869 in control cells to 55,388 following SFPQ knockdown (Supplementary Fig. 4a–c), suggesting dysregulated chromatin incorporation of H3.3. Despite the increase in H3.3 peak numbers following SFPQ depletion, we observed marked log₂-fold reductions in H3.3 peak height across all annotated sequence categories (Fig. 4C). A genome-wide area under the curve (AUC) analysis of log₂-fold changes further confirmed a global decrease in H3.3 chromatin occupancy in SFPQ-deficient cells (Supplementary Fig. 4a, f). These results support a model in which disruption of SFPQ–DAXX function redirects H3.3 deposition from discrete high-occupancy loci to a more diffuse, moderate-occupancy pattern across the genome. To validate the ChIP–seq findings, we performed ChIP–PCR on representative genomic regions previously shown to accumulate R-loops upon SFPQ loss (Fig. 1E). Consistent with impaired recruitment, SFPQ depletion led to reduced DAXX abundance at α -satellite, SatII, SatIII, (sub)telomeric regions, LINE1, AluY, and AluS elements (Fig. 4D, E). This redistribution was accompanied by a significant reduction in histone H3.3 abundance at affected loci, without changes in canonical histone H3 density (Fig. 4D, E). These effects were recapitulated at the promoters of *PD3A* and *SYCP2*, previously identified as SFPQ and DAXX target genes, respectively (Fig. 4F)^{56,61}. In contrast, the R-loop-resistant *WRNIP* promoter exhibited no changes in chromatin marks following SFPQ depletion (Fig. 1F; Supplementary Fig. 1k, 4g). Furthermore, the *SSH* promoter, which lacks SFPQ binding, showed no detectable DAXX or H3.3 occupancy, reinforcing the specificity of our ChIP assays (Supplementary Fig. 4g). Importantly, SFPQ depletion did not affect total histone H3.3 protein levels, ruling out limited H3.3 availability as a contributing factor to the observed chromatin changes

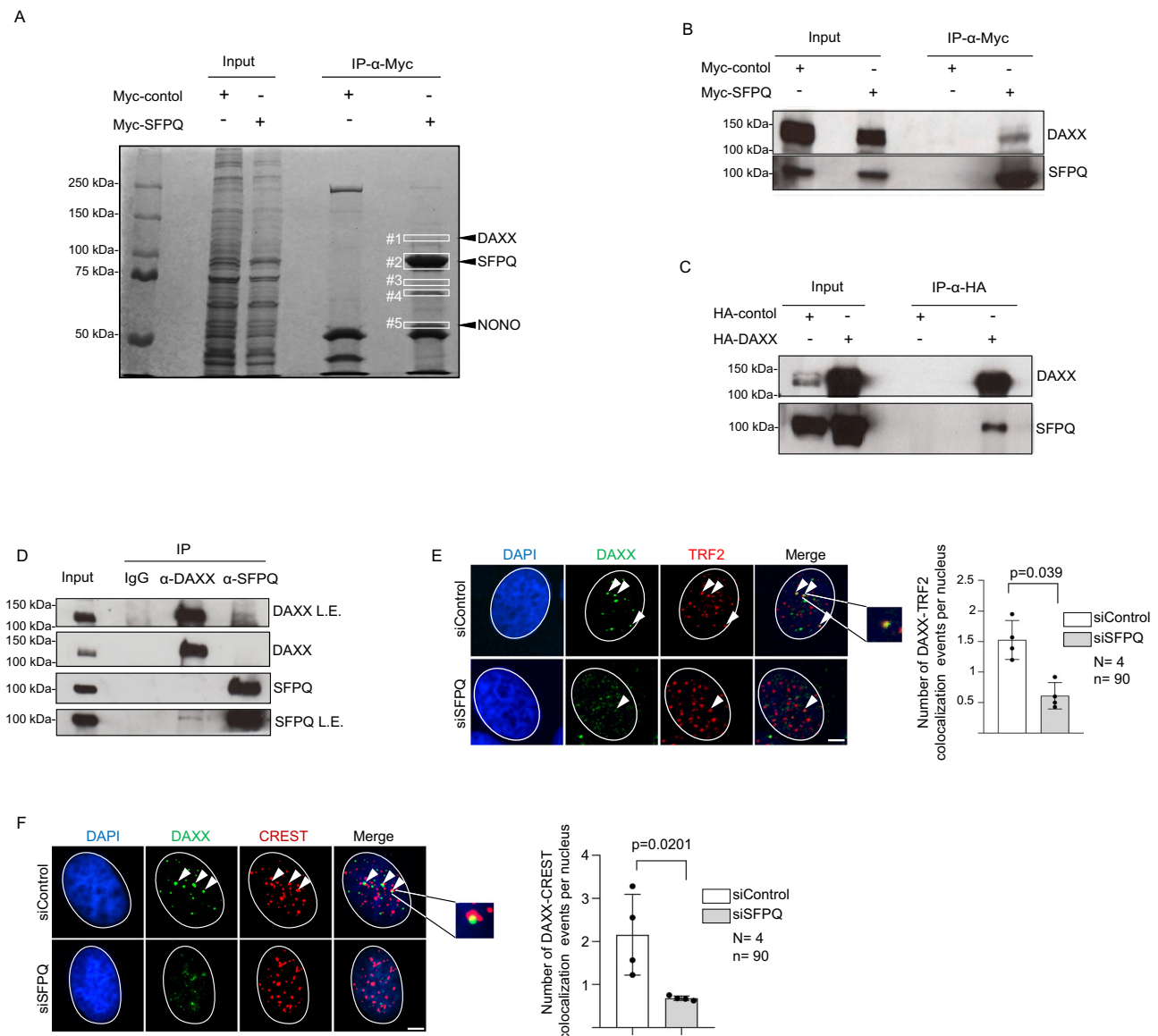


Fig. 3 | SFPQ interacts with the histone H3.3 chaperone DAXX. **A** Anti-SFPQ immunoprecipitation of ectopically expressed myc-tagged SFPQ followed by SDS-PAGE, Coomassie staining and identification of interacting proteins by MALDI-TOF mass spectrometry. **B** Immunoprecipitation of overexpressed myc-tagged SFPQ in H1299 cells. **C** Immunoprecipitation of overexpressed HA-DAXX in H1299 cells. **D** Co-immunoprecipitation of endogenous SFPQ and DAXX proteins in U-2 OS cells using specific antibodies. L.E., low exposure. **E** Combined immunofluorescence using anti-DAXX and anti-TRF2 antibodies in U-2 OS cells transfected with indicated siRNAs. Right panel, quantification of DAXX-TRF2 co-localization events. **F** Representative images of combined immunofluorescence with anti-DAXX and

anti-CREST (centromere) antibodies in U-2 OS cells transfected with indicated siRNAs. Right panel, quantification of DAXX-CREST colocalization events. **A–D** representative images are shown. For quantifications, data represents the mean, error bars indicate standard deviation. N = number of independent experiments. n = number of analyzed nuclei. Arrowheads and zoomed images indicate co-localization events. An unpaired, two-sided Student's t test was used to calculate statistical significance; p values are shown. Scale bar (1 μ m) applies to all images in respective immunofluorescence panels. DNA was stained using DAPI (4',6-diamidino-2-phenylindole). Source data are provided with this paper.

(Supplementary Fig. 4h). Together, these findings support a model in which SFPQ binds R-loops and facilitates DAXX-dependent deposition of histone H3.3 at repetitive genomic elements.

The Proline-rich domain and RRM motifs of SFPQ are central for R-loop suppression

To dissect the molecular basis of the SFPQ–DAXX interaction, we transiently expressed full-length HA-tagged DAXX together with a panel of myc-tagged SFPQ deletion mutants lacking defined protein domains, followed by co-immunoprecipitation assays (Fig. 5A; Supplementary Fig. 5a). Notably, deletion of the intrinsically disordered proline-rich domain (residues 105–204) abolished the interaction with

full-length DAXX (Fig. 5A, B; SFPQ Δ P). Other SFPQ domains did not significantly contribute to DAXX binding (Supplementary Fig. 5a, b). By contrast, the proline-rich P-domain, when fused to GFP and a nuclear localization signal, was sufficient to interact with full-length DAXX (Fig. 5C; Supplementary Fig. 5c). The requirement for the P-domain was further validated in pull-down assays using recombinant DAXX and recombinant full-length or Δ P SFPQ proteins (Fig. 5D). Consistent with this, whole-exome sequencing of an osteosarcoma patient cohort identified a P-domain mutation involving a non-frameshift substitution of Tyr150 with Lys-Pro (T150 Δ PK)⁶². Full-length SFPQ carrying this mutation (SFPQ nFS) displayed impaired binding to ectopically expressed HA-DAXX in co-immunoprecipitation assays (Fig. 5B, C;

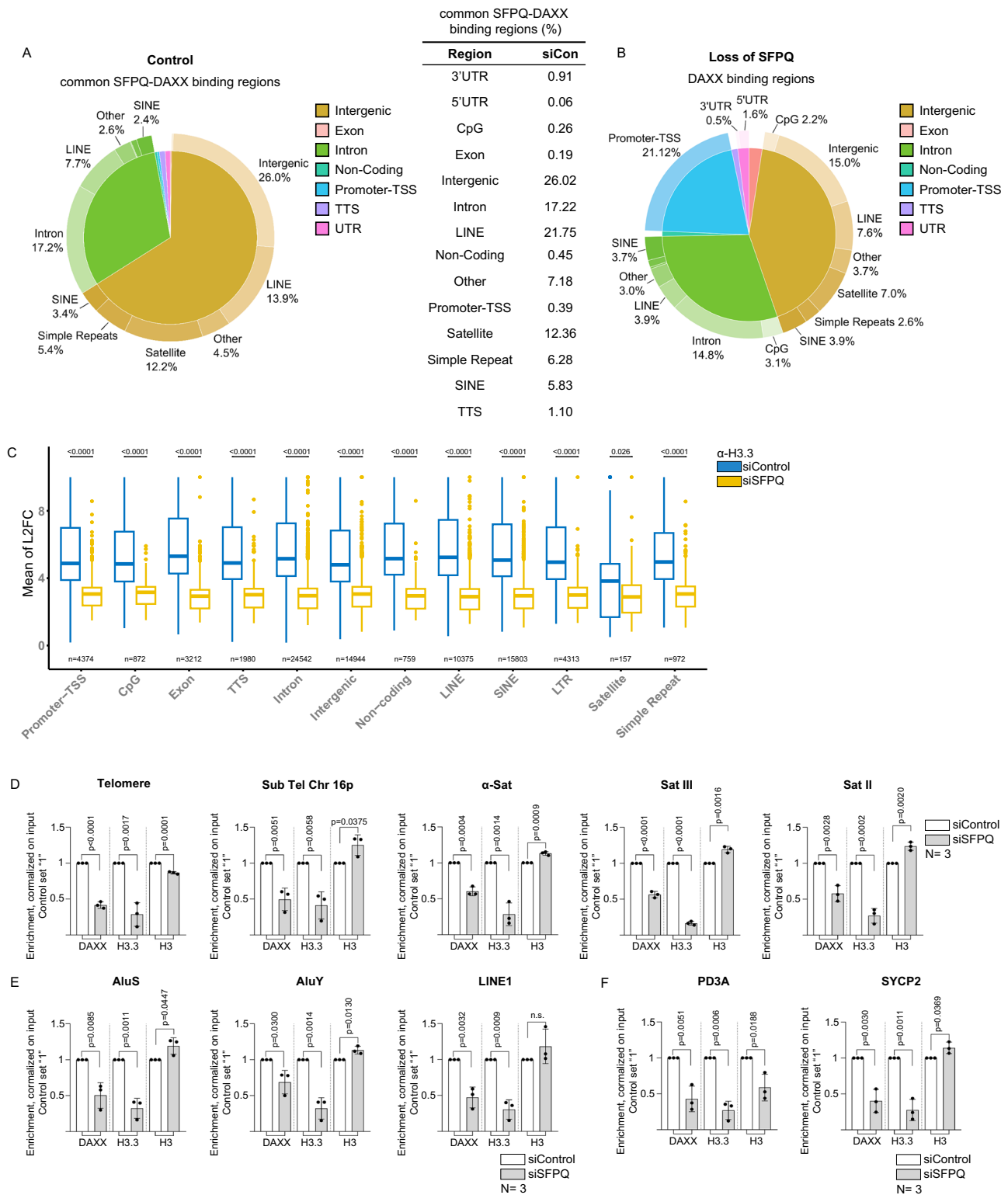
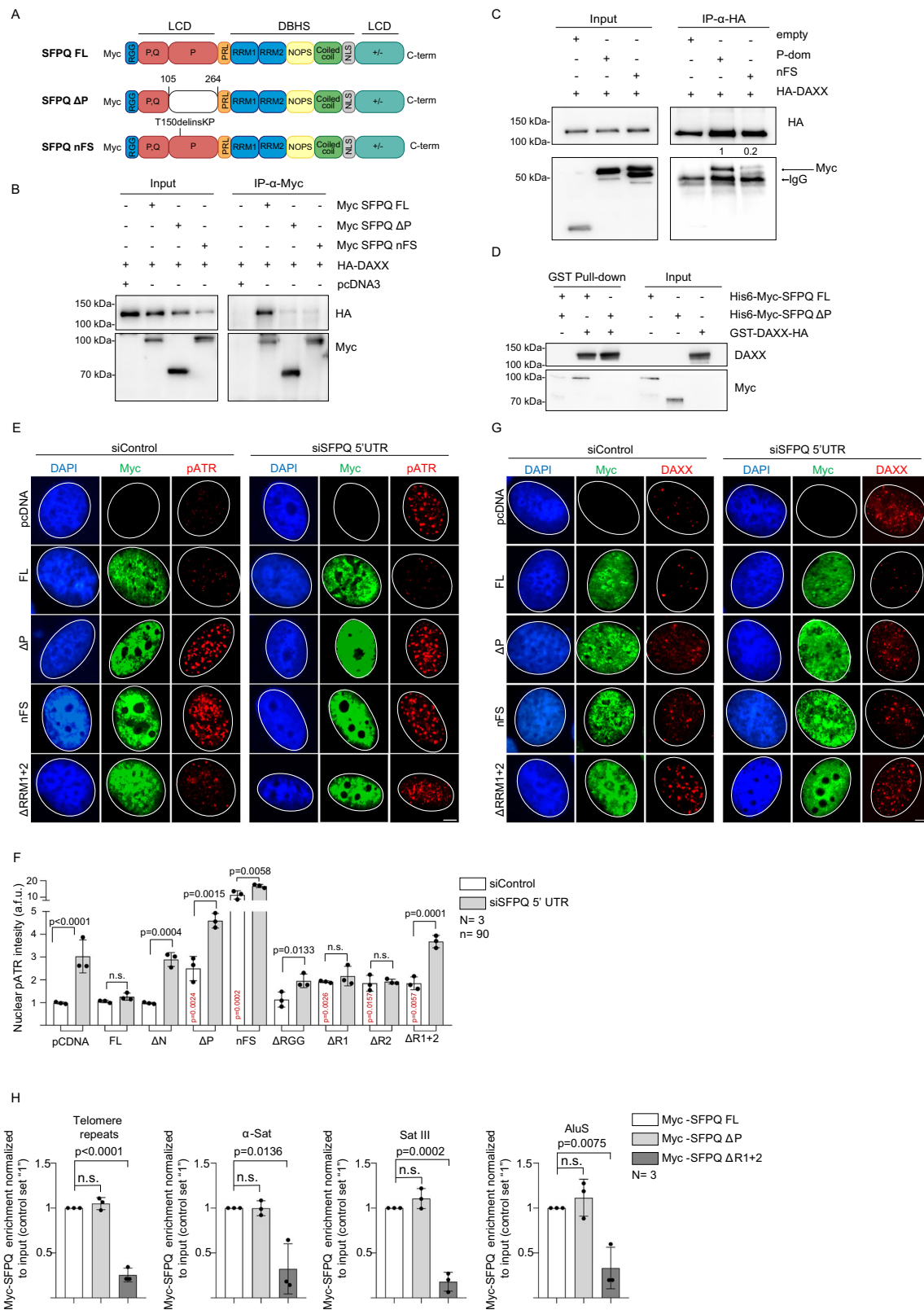


Fig. 4 | SFPQ controls DAXX localization and deposition of histone H3.3. **A** Pie chart representing percentage of SFPQ-DAXX common binding sites across indicated genome regions in U-2 OS cells transfected with control siRNAs (Control). Right panel, common SFPQ-DAXX peaks in control cells. **B** Pie chart representing the percentage of DAXX peaks in indicated genome regions in SFPQ depleted U-2 OS cells (loss of SFPQ). **C** Box-plot representing H3.3 mean Log₂ fold change at reported genomic regions in control (blue) and SFPQ depleted (yellow) U-2 OS cells. *N*, number of peaks for reported genomic region for both conditions. *P*-values were calculated with Wilcoxon test and adjusted for False Discovery Rate. Center, median (50th percentile); box bounds, Q1 (25th percentile) and Q3 (75th percentile); whiskers, minimum and maximum non-outlier values (within 1.5 × IQR of Q1/

Q3); dots, outliers. Promoter-TSS Promoter-Transcription Starting Site, CpG CpG islands, TTS Transcription Termination Site, LINE Long Interspersed Nuclear Element, SINE Short Interspersed Nuclear Elements, LTR Long Terminal Repeat. Chromatin immunoprecipitation (ChIP) assay performed on U-2 OS cells transfected with indicated siRNAs. Data for (peri)centric and (sub)telomeric repeats (**D**) transposable elements (**E**) and positive control regions for SFPQ and DAXX are shown. For quantifications shown in (**D-F**), data represents the mean values, error bars indicate standard deviation. *N* = 3. An unpaired, two-sided Student's *t* test was used to calculate statistical significance; *p*-values are shown. Source data are provided with this paper.



Supplementary Fig. 5c). We next examined the contribution of individual SFPQ domains to DAXX localization and the suppression of R-loop-induced replication stress. To this end, we generated U-2 OS cell lines stably expressing myc-tagged SFPQ deletion mutants lacking the 5' and 3' UTRs (Fig. 5A; Supplementary Fig. 5a). Endogenous full-length SFPQ was selectively depleted using 5'UTR-specific siRNAs, allowing assessment of truncated SFPQ variants on DAXX distribution and

replication stress (Supplementary Fig. 5d). In control cells stably transfected with a pcDNA control vector and subsequently treated with 5'UTR-targeting siRNAs, DAXX became delocalized and p-ATR (Thr1989) foci emerged, consistent with replication stress (Fig. 5E, F). As expected, ectopic expression of full-length myc-tagged SFPQ rescued phenotypes associated with depletion of endogenous SFPQ (Fig. 5E, F). In contrast, ectopic expression of myc-tagged ΔP or nFS

Fig. 5 | Specific SFPQ domains mediate DAXX recruitment and R-loop binding. **A** SFPQ protein domains. Full-length (FL) SFPQ is composed by the arginine-glycine-glycine (RGG) box (blue); N- and C-terminal low-complexity domains (LCDs); proline/glutamine-rich subdomain (P,Q, red); proline-rich domain (P, red); PR linker (PRL, orange); DBHS conserved region (DBHS) containing RNA-recognition motifs (RRM1 and RRM2, blue), NonA/paraspeckles (NOPS) domain (yellow), the coiled-coil domain (green), a nuclear-localization sequence (NLS) (gray); and the +/- domain (bright green). Deleted domains are reported as white boxes; number refer to amino acid positions; patient mutation in nFS construct is indicated. **B** Co-immunoprecipitation of myc-tagged SFPQ versions and HA-tagged DAXX in U-2 OS cells. **C** Co-immunoprecipitation of myc-tagged SFPQ P-domain versions (myc-P-dom or myc-nFS) and HA-DAXX in U-2 OS cells. Arrowheads indicate the position of immunoglobulin bands; numbers indicate immunoprecipitation efficiency. **D** Pull-down assay of recombinant GST-DAXX-HA with recombinant His₆-myc-SFPQ FL or His₆-myc-SFPQ ΔP using anti-GST beads. **E** Representative images of combined immunofluorescence with anti-myc and anti-p-ATR(Thr1989) antibodies after transient expression of wild-type and mutant SFPQ versions

lacking the SFPQ 5'UTR and 3'UTR sequences. Endogenous SFPQ was knocked down using 5'UTR-specific siRNAs. Myc-positive cells were considered for quantification. **F** Quantification of pan-nuclear p-ATR in myc-positive cells described in (E). **G** Representative images of combined immunofluorescence with anti-myc and anti-DAXX antibodies in U-2 OS cells stably expressing SFPQ mutants. Endogenous SFPQ was knocked down using 5'UTR-specific siRNAs. **H** Anti-myc ChIP-qPCR analysis on U-2 OS cells transfected with indicated wild-type and mutant SFPQ expression constructs. Endogenous SFPQ was knocked down using 5'UTR-specific siRNAs. **B–D** representative images are shown. In quantifications, data represents the mean; error bars, standard deviation. *N* = number of independent experiments; *n* = analyzed nuclei. Statistical test, unpaired, two-sided Student's *t* test. Scale bar (1 μm) applies to all images in respective immunofluorescence panels. DNA was stained using DAPI (4',6-diamidino-2-phenylindole). **F**; *p*-values (in black) refer to control siRNA and siSFPQ-transfected cells; *p*-values (in red) refer to U-2 OS cells stably transfected with a control pcDNA versus cells expressing mutant SFPQ versions and were transfected with control siRNAs. Source data are provided with this paper.

SFPQ variants resulted in elevated basal replication stress, which was further exacerbated upon siRNA-mediated knockdown of endogenous SFPQ, consistent with a dominant-negative effect of the mutant proteins (Fig. 5E, F). Notably, a subset of p-ATR foci induced by ΔP or nFS SFPQ colocalized with telomeres (Supplementary Fig. 5e). This is in agreement with the role of wild-type SFPQ in suppressing R-loop-associated replication stress at telomeres⁴⁶. Replication stress induced by ΔP or nFS SFPQ variants was accompanied by DAXX delocalization (Fig. 5G; Supplementary Fig. 5g). Similarly, expression of myc-tagged SFPQ lacking the RNA-binding motifs (ΔRRM1 + 2) elevated basal replication stress, which was further exacerbated by siRNA-mediated depletion of endogenous SFPQ and likewise coincided with DAXX delocalization (Fig. 5E, F; Supplementary Fig. 5f, g). U-2 OS cells expressing single ΔRRM1 or ΔRRM2 SFPQ deletion constructs exhibited a modest but significant increase in replication stress, which did not further intensify upon depletion of endogenous SFPQ. By contrast, ΔRG SFPQ expression did not elevate basal replication stress and induced only minor changes in p-ATR (Thr1989) foci following loss of endogenous SFPQ (Fig. 5E, F; Supplementary Fig. 5f). To assess the role of the P, RRM1 and RRM2 domains in SFPQ recruitment to chromatin, we transiently transfected U-2 OS cells with control, myc-tagged full-length, ΔP or ΔRRM1 + 2 SFPQ constructs while depleting endogenous SFPQ. Anti-myc ChIP assays revealed that the ΔRRM1 + 2 variant failed to localize to R-loop-rich regions, including telomeres, α-satellite, SatIII, SatII, and AluS sequences (Fig. 5H). By contrast, the P-domain was dispensable for chromatin loading (Fig. 5H). Together, these findings establish RRM1 and RRM2 as essential for SFPQ recruitment to R-loop hotspots, whereas the P-domain is specifically required for DAXX recruitment.

Loss of SFPQ drives R-loop-mediated genome instability and mitotic defects

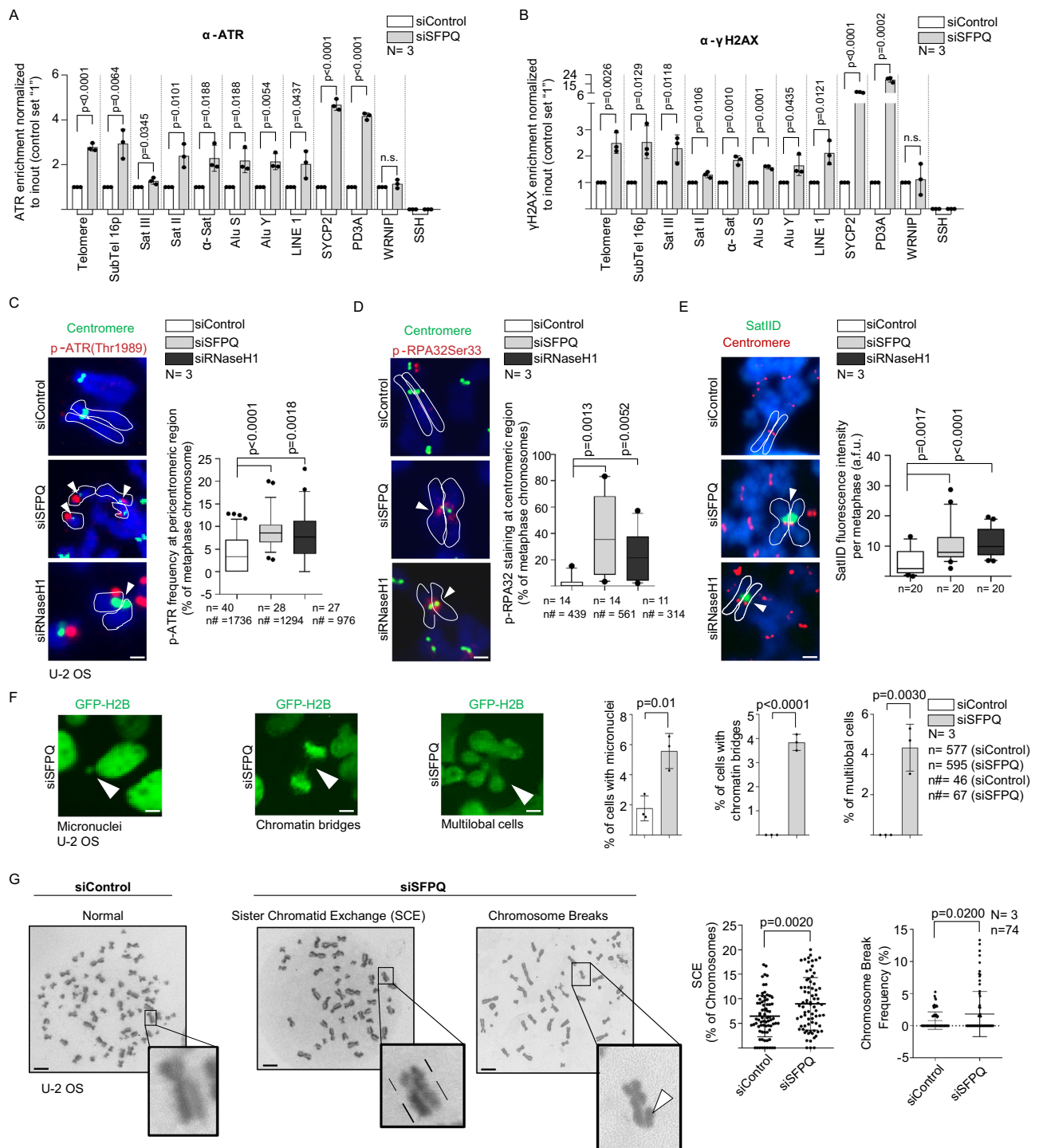
We next assessed the downstream consequences of unprogrammed R-loops by performing ChIP analysis in SFPQ-depleted U-2 OS cells. Loss of SFPQ from chromatin resulted in increased accumulation of ATR, RPA32 and γH2AX at telomeres, subtelomeres, α-satellite, SatII, SatIII and SINE/LINE repeats, underscoring the central role of SFPQ in suppressing replication stress and DNA damage at repetitive regions (Fig. 6A, B; Supplementary Fig. 6a, d). This effect was accompanied by recruitment of FANCD2 and RAD51, indicative of local activation of DNA repair pathways that counteract genomic instability (Supplementary Fig. 6b, c). Phenotypes associated with loss of SFPQ function were reproduced in independent H1299 cells (Supplementary Fig. 6e). Notably, a recent study showed that aberrant R-loop regulation at centromeres leads to mitotic defects, providing a compelling model to interrogate the role of SFPQ in safeguarding genome stability⁶³.

We analyzed R-loop formation and replication stress at centric and pericentric repeats using native metaphase chromosome spreads. Loss of SFPQ resulted in pronounced accumulation of activated ATR (pATR-Thr1989) and increased phosphorylation of RPA32 at Ser33 (pRPA32Ser33) at centromeres, indicative of replication stress (Fig. 6C, D). This effect was accompanied by accumulation of pericentric SatIII transcripts at centromeres of selected chromosomes, as detected by RNA-FISH (Fig. 6E). Phenotypes observed in SFPQ-deficient cells were reproduced in RNaseH1 knockdown cells (Fig. 6C–E). To further investigate mitotic progression, we performed time-lapse microscopy following transient SFPQ depletion in U-2 OS and H1299 cells stably expressing GFP-tagged histone H2B. Loss of SFPQ markedly increased the incidence of chromatin bridges, multilobular cells and micronuclei (Fig. 6F; Supplementary Fig. 6f; Supplementary Movie 1). To assess the impact of R-loops on chromosome stability, we next analyzed mitotic chromosome defects using Giemsa-stained metaphase spreads. U-2 OS and MCF-10A cells were transiently depleted of SFPQ and allowed to incorporate BrdU during two rounds of semi-conservative DNA replication. Metaphase spreads were subsequently UV-irradiated prior to Giemsa staining, enabling discrimination of sister chromatids by differential staining intensity. Loss of SFPQ increased chromatid breaks and reciprocal sister chromatid exchanges (SCEs) in both cell models (Fig. 6G; Supplementary Fig. 6g). As expected, these phenotypes were accompanied by alterations in cell-cycle progression (Supplementary Fig. 1g; Supplementary Fig. 6h). Collectively, these findings establish SFPQ as a key suppressor of R-loop-mediated genome instability.

SFPQ-dependent R-loop resolution suppresses innate immunity pathways in human sarcoma

To investigate the physiological consequences of SFPQ loss of function, we performed RNA sequencing (RNA-seq) in U-2 OS cells subjected to SFPQ knock-down. Principal component analysis revealed marked alterations in the transcriptomic landscape (Supplementary Fig. 7a). Differential expression analysis identified 1268 genes significantly upregulated and 1163 genes downregulated upon SFPQ depletion (Supplementary Fig. 7b, c). Gene ontology enrichment demonstrated a pronounced activation of pathways associated with cytokine production and signaling, immune responses, antiviral defense, and innate immunity (Fig. 7A; Supplementary data 7).

Network analysis revealed activation of gene expression programs associated with epidermal development and secretion, alongside a 29-gene signature linked to innate immunity encompassing antiviral responses, cGAS–STING-mediated interferon activation, and inflammatory pathways (Fig. 7B). Enhanced interferon signaling provides a mechanistic basis for the elevated DAXX expression observed in SFPQ-depleted cells (Supplementary Fig. 3h, I)⁶⁴. The presence of cytoplasmic DNA species indicates activation of the cGAS–STING axis,



culminating in IRF3 and NF- κ B transcriptional activity. This cascade drives interferon responses and the production of pro-inflammatory cytokines, ultimately reshaping the tumor microenvironment^{65,66}. Independent quantitative RT-PCR and western blot analyses confirmed the upregulation of canonical target genes associated with the cGAS-STING axis, interferon signaling, inflammation, and NF- κ B activation in both, U-2 OS and H1299 cells following *SFPQ* knock-down (Fig. 7C, D; Supplementary Fig. 7d, e). These findings underscore a broad role for SFPQ in restraining innate immune pathways. Immunofluorescence revealed that depletion of SFPQ or DAXX increased the frequency of micronuclei, which co-stained with cGAS, consistent with the exposure of cytoplasmic DNA (Fig. 7E). The presence of cytoplasmic DNA in

SFPQ-deficient U-2 OS cells was further validated by immunofluorescence using anti-dsDNA antibodies (Supplementary Fig. 7f). Transient depletion of SFPQ in U-2 OS and H1299 cells resulted in robust phosphorylation of the transcription factor IRF3, a key driver of interferon-stimulated gene expression (Fig. 7F; Supplementary Fig. 7g). Immuno-DNA FISH analysis further demonstrated that SFPQ loss increased the frequency of micronuclei stained with CREST antibodies, pericentric SatIID probes, or telomeric repeat probes, suggesting that enhanced R-loop-mediated genome instability at these regions may contribute to micronuclei formation (Supplementary Fig. 7h). To directly assess the role of SFPQ in micronuclei formation, we utilized U-2 OS cells ectopically expressing SFPQ deletion

Fig. 6 | Loss of SFPQ mediates defects in mitotic progression. ChIP on U-2 OS cells transfected with control and SFPQ-specific siRNAs. The graphs report the binding of ATR (A) and γ H2AX (B) to the indicated region of interest. Representative images of native metaphase spreads obtained from control and SFPQ knock-down U-2 OS silenced with indicated siRNAs. Immunostaining was carried out using anti-p-ATR(Thr1989) and anti-CREST (centromere) antibodies (C) or anti-p-RPA32(Ser33) and anti-CREST antibodies (D). Left panels, representative images. Arrowheads indicate position of p-RPA32Ser33/p-ATR staining in vicinity to centromeres. Right panels, quantification of positive staining at (peri)-centromeric regions. E Immuno-RNA FISH staining of native metaphase chromosomes derived on siRNA-transfected U-2 OS cells. Staining with anti-CREST antibody and a pericentric SatIID RNA-FISH probe is shown. Left panel, representative images. Right panel, quantification of SatIID signals at (peri)centric regions; Arrowheads, position of SatIID RNA-FISH staining in vicinity to centromeres. F Representative images of time-lapse microscopy using U-2 OS cells constitutively expressing GFP-tagged histone H2B after siRNA transfection. Left panels, representative images of

observed mitotic defects. Right panels, relative quantifications of mitotic defects. G Giemsa-stained metaphases of U-2 OS cells transfected with indicated siRNAs to evaluate rates of sister chromatid exchange (SCE) and chromosome breaks. Left panels, representative images. Right panels, quantification of metaphase chromosome defects. Zoomed images, observed defects. Bars in data blots represent mean values, error bars indicate standard deviation. N = number of independent experiments. An unpaired, two-sided Student's t test was used to calculate statistical significance; p values are shown. For (C–E), n = number of analyzed metaphase spreads, $n\#$ = number of analyzed chromosomes, graph shows median and 10–90 percentile; dots, outliers. Scale bars correspond to (C–E), 0.1 μ m; (F) 1 μ m and (G) 5 μ m. Scale bars apply to all images in the respective immunofluorescence panels. C–E DNA (blue) was stained using DAPI (4',6-diamidino-2-phenylindole). For (F), n = number of analyzed nucleus (used for micronuclei quantification), $n\#$ = number of analyzed cell divisions (used for chromatin bridges and multilobed cells qualifications). For G, n = number of analyzed metaphases. Source data are provided with this paper.

constructs. Consistent with data on replication stress and chromatin recruitment, ectopic expression of myc-tagged Δ P or Δ RRM1 + RRM2 variants failed to rescue elevated micronuclei frequency observed in cells lacking endogenous SFPQ (Fig. 7G). Strikingly, expression of the patient-derived SFPQ nFS variant exerted a dominant-negative effect, driving micronuclei formation, which was further exacerbated upon depletion of endogenous SFPQ (Fig. 7G). These findings highlight the importance of the SFPQ P and RRM1 + RRM2 domains in maintaining genome stability. To directly test whether elevated R-loop levels underlie micronuclei and cytoplasmic DNA formation in SFPQ-deficient cells, we employed U-2 OS cells carrying a doxycycline-inducible mCherry-tagged RNaseH1 (ref. 67). Ectopic RNaseH1 expression suppressed both micronuclei formation in SFPQ- or DAXX-depleted cells and the activation of 10 out of 11 pro-inflammatory genes (Fig. 7H, I; Supplementary Fig. 7i). Notably, consistent with the reported role of SFPQ as a transcriptional repressor of IL8 (ref. 48) SFPQ depletion increased *IL8* expression in an RNaseH1-independent manner (Supplementary Fig. 7i). Although U-2 OS cells have been reported to express STING at low or undetectable levels (refs. 68–70), SFPQ knock-down induced robust STING phosphorylation. Moreover, treatment with the STING antagonist H151 suppressed the expression of pro-inflammatory genes, including *IFIT1*, *INFB1*, *CXCL10*, *IL1-alpha*, *IRF2*, and *IL12*, while leaving *CGAS*, *STING*, and *IL8* mRNA levels unaffected (Supplementary Fig. 7j, k).

To explore the clinical relevance of SFPQ-mediated regulation of innate immunity in human cancer, we analyzed patient survival data. Across pan-cancer cohorts, SFPQ expression showed no significant association with overall survival (Supplementary Fig. 7l). However, selective analysis of the TCGA sarcoma dataset revealed that elevated SFPQ expression correlated with reduced patient survival (Fig. 7j). In contrast, high expression of the innate immunity gene signature identified by RNA-seq in U-2 OS cells lacking SFPQ (29 genes including *IFIT1*, *ACTA2*, *UNC13D*, *IFI44*, *STING1*, *OASL*, *MX2*, *NLRP1*, *MLKL*, *RNF125*, *APOBEC3H*, *RTP4*, *PLA2G10*, *BIRC3*, *ABCC9*, *MMP12*, *DDX60*, *OAS2*, *PTPN22*, *IFIT2*, *IFIT3*, *AIM2*, *CCL5*, *CXCL10*, *RSAD2*, *HERC5*, *TSPAN32*, *ZBP1*, and *IL1A*) was associated with improved overall survival (Fig. 7B, K; Supplementary Fig. 8; Supplementary Table 1).

Consistent with these findings, SFPQ expression inversely correlated with the innate immunity gene signature in human sarcoma (Fig. 7L). Patients with low SFPQ and high innate immunity signature expression exhibited markedly improved overall survival compared to those with high SFPQ and low signature expression (Fig. 7M; Supplementary Fig. 9; Supplementary Table 2). Together with mechanistic evidence from cell lines, these data support a central role for SFPQ in suppressing R-loop-mediated genome instability at repetitive sequences, thereby preventing activation of innate immune pathways via the cGAS-STING axis and fostering a tumor-promoting

microenvironment in sarcoma. Accordingly, the SFPQ-based gene expression signature provides a valuable tool for improved patient stratification and highlights the SFPQ–DAXX interaction as a potential therapeutic target to stimulate antitumor immunity.

Discussion

Recent studies have shown that transcription of telomeric and centromeric regions, simple/low-complexity repeats, satellite repeats, rDNA arrays, and retroelements can promote R-loop formation^{25,26,28,34}. We demonstrate that SFPQ safeguards genome stability by preventing R-loop accumulation across major classes of repetitive elements in the human genome. Loss of SFPQ led to pronounced R-loop accumulation at telomeres, subtelomeres, α -satellite, SatII, SatIII, LINE1, AluS, and AluY elements. Ectopic induction of R-loops triggered SFPQ recruitment to these repeat regions, indicating a dynamic role for SFPQ in sensing R-loop abundance. In vitro, SFPQ bound synthetic R-loops with protruding RNA termini but failed to interact with RNA:DNA duplexes, revealing a preference for three-stranded nucleic acid structures.

A comparable preference for R-loop structures over RNA:DNA duplexes has recently been reported for the Tudor domain-containing protein TDRD3, which recruits the RNA-dependent helicase DHX9 to resolve R-loops at gene promoters, the FANCI-FANCD2 complex, and ATRX^{71–73}. We found that the RNA-recognition motifs RRM1 and RRM2 of SFPQ are required for its binding to R-loop-containing chromatin in U-2 OS cells. Consistent with our hypothesis that SFPQ functions as a recruitment factor for R-loop management machineries, we identified the replication-independent H3.3 histone chaperone DAXX as a novel SFPQ-interacting protein. DAXX has been reported to interact with the ATP-dependent helicase ATRX to incorporate the replication-independent histone variant H3.3 into nucleosomes at repetitive elements, including telomeres, thereby maintaining repressive chromatin structure^{37,59,74–88}. Independent studies have implicated ATRX and DAXX in suppressing R-loops, although the mechanism of complex recruitment has remained unclear^{13,39}. We show that DAXX binds to the proline-rich (P) domain of SFPQ; accordingly, expression of an SFPQ P-domain deletion construct recapitulated the replication stress and aberrant DAXX localization observed in SFPQ-depleted U-2 OS cells. These findings support a model in which SFPQ employs its RRM1 and RRM2 domains to bind R-loops at repetitive elements and subsequently recruits DAXX-dependent histone H3.3 chaperone activity via its P-domain to re-establish functional nucleosome templates. Notably, SFPQ depletion in HI299 cells was accompanied by reduced ATRX protein, but not mRNA, levels—recapitulating phenotypes observed in DAXX-deficient mouse embryonic stem cells⁵⁹. Together, these results underscore the central importance of SFPQ in the assembly of a functional DAXX–ATR complex. In line with the role of SFPQ in recruiting DAXX, CHIP-seq analysis revealed that loss of SFPQ causes

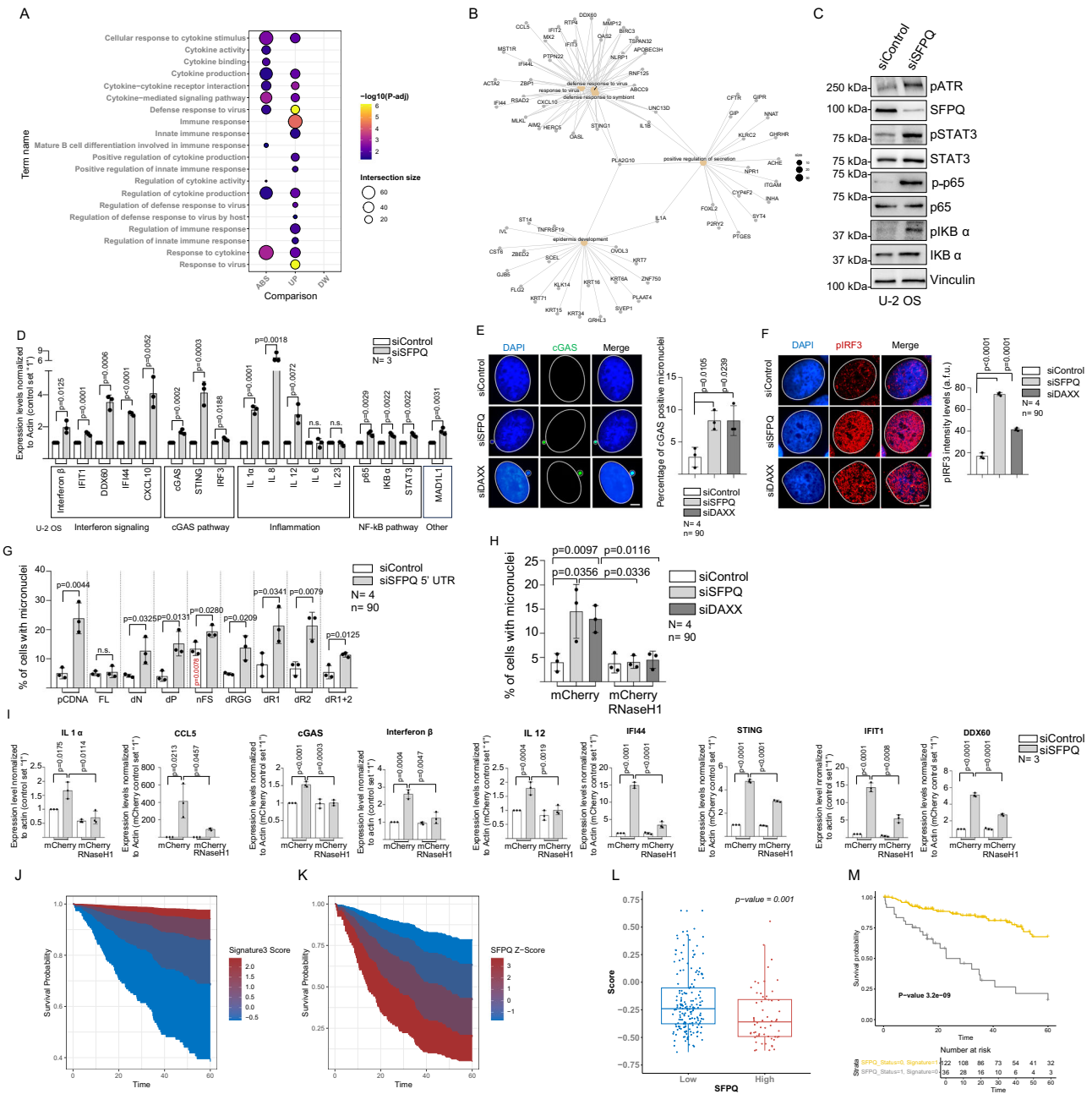


Fig. 7 | Loss of SFPQ mediates activation of pathways of innate immunity.

A Enriched terms in SFPQ knock-down U-2 OS cells. UP, upregulated; DW, down-regulated; ABS, absolute value terms. Circle size, number of genes; color code, $-\log_{10}$ alterations (p -adjusted); enrichment p -values were computed using a one-sided Fisher's exact and Benjamini-Hochberg FDR correction. **B** Network representation for 5 enriched biological terms in siSFPQ U-2 OS cells. Orange nodes, term names; black dots, individual genes. **c** Western blot of siRNA-transfected U-2 OS cells. **D** Quantitative RT-PCR of indicated genes in siSFPQ U-2 OS cells. **E** cGAS immunofluorescence analysis. Left panel, representative images. Right panel, percentage of micronuclei with cGAS staining. **F** Immunofluorescence on siRNA-transfected U-2 OS cells. Left panel, representative images. Right panel, quantification of pan-nuclear pIRF3 staining. **G** Quantification of micronuclei in U-2 OS cells stably expressing my-tagged, full-length or mutant SFPQ. Endogenous SFPQ was depleted using 5'UTR-specific siRNAs. **H** Micronuclei in U-2 OS cells after transfection with indicated siRNAs and expression of mCherry-tagged RNaseH1 or control-mCherry. **I** Quantitative RT-PCR analysis in siRNA-transfected U-2 OS cells

expressing mCherry-tagged RNaseH1 or control-mCherry. Survival plot showing correlation between low SFPQ Z-score and improved patient overall survival (**j**) high signature score (Fig. 7b) and improved patient overall survival (**K**). **L** Correlation of expression between SFPQ and pro-inflammatory gene signature. Score based on mean Z-score trimmed at 20%. The p -value was obtained by Anova-test. Center, median (50th percentile); box bounds, Q1 (25th percentile), Q3 (75th percentile); whiskers, minimum and maximum non-outlier values; dots, individual samples. **M** Kaplan-Meier plot analysis based on innate immunity signature expression and survival. Yellow, low SFPQ expression (Status=0) and high signature expression (Signature=1); gray, high SFPQ expression (Status=1) and low signature expression (Signature=0). Patient numbers are shown. P -value was calculated using a two-sided log-rank (Mantel-Cox) test. **D-I** Data presented as mean values \pm standard deviation. N , number of independent experiments; n , analyzed nuclei; statistical test, unpaired, two-sided Student's t test; scale bar (5 μ m) applies to all images; DAPI, 4',6-diamidino-2-phenylindole. **J-M** patient data were obtained from the TCGA pancreatic sarcoma dataset. Source data are provided with this paper.

genome-wide DAXX delocalization and a global reduction in histone H3.3 abundance across all tested repeat elements. Thus, SFPQ depletion recapitulates the impaired H3.3 deposition at repetitive elements previously observed in model systems lacking DAXX or ATRX expression^{37,59,74,77,79,80,82,84,86}. Notably, studies in mouse embryonic stem cells have shown that ATRX binding to trimethylated histone H3 lysine 9 (H3K9me3) recruits DAXX to endogenous retroviral elements (ERVs)^{59,89}. Interestingly, DAXX has been shown to interact with a SETDB1-KAP1-HDAC1 protein complex in an ATRX-independent manner⁹⁰. KAP1, together with the nucleosome remodeler SMARCAD1, directs SETDB1 deposition at ERVs to enforce H3K9me3 and HP1-mediated repression; in addition, SMARCAD1 has been reported to influence H3.3 incorporation^{91–94}. Future studies should clarify the role of R-loops and SFPQ in coordinating ATRX-dependent and ATRX-independent functions of DAXX in mediating H3.3 deposition and epigenetic silencing through KAP1, SETDB1, and SMARCAD1. R-loop-mediated replication stress at repetitive elements in SFPQ-depleted cells led to the recruitment of R-loop-associated DNA damage repair factors, including RAD51 and FANCD2^{72,95–97}. Failure of the cellular response to ectopic R-loop induction resulted in DNA breaks, elevated sister chromatid exchange (SCE) events, and defects in mitotic progression, accompanied by the accumulation of SatIID transcripts and replication stress markers at (peri)centric regions of native metaphase chromosomes. These findings are consistent with a recent study reporting mitotic instability driven by excessive R-loop formation and ATR activity at centromere regions⁶³. In addition, multiple R-loop regulators—including BRCA1, BRCA2, RBMX, DHX38, SETX, FUS, DAXX, FA pathway components, and the BLM–TOP3A–RMI1–RMI2 (BTR) complex—have been implicated in maintaining (peri)centromere stability (for review ref. 25). Genome instability in cancer cells promotes the accumulation of cytoplasmic DNA species, including micronuclei, which activate the cGAS–STING pathway and drive the expression of pro-inflammatory cytokines and type I interferons, thereby stimulating immune cells and enhancing antigen presentation within the tumor microenvironment^{65,66}. Consistently, mitotic instability induced by SFPQ depletion resulted in cytoplasmic dsDNA accumulation and increased micronuclei frequency, recapitulating phenotypes previously observed in cells lacking DAXX, ATRX, or H3.3^{60,87,98,99}. Micronuclei arising in SFPQ-depleted cells matured into cytoplasmic DNA, as evidenced by cGAS recruitment, leading to full activation of the cGAS–STING pathway via NF- κ B and IRF3 and induction of key innate immunity genes. Ectopic expression of RNaseH1 suppressed micronuclei formation and blocked activation of both cGAS–STING signaling and innate immunity gene expression. Consistently, treatment with the STING antagonist H151 recapitulated the suppression of innate immune gene expression. Thus, R-loop-mediated genome instability represents a primary source of cytoplasmic DNA production and cGAS–STING pathway activation in SFPQ loss-of-function cells. Nevertheless, we cannot exclude contributions from repeat RNAs escaping R-loop structures at DNA repeats, cytoplasmic R-loops, or chromatin alterations at promoters linked to interferon response genes^{100,101}. Data from osteosarcoma whole-exome sequencing (WES) and TCGA sarcoma cohorts underscore the clinical relevance of SFPQ in maintaining R-loop-mediated genome stability. A recent WES study of 31 osteosarcoma patients identified two disabling SFPQ mutations⁶². Expression of a Tyrosine150-to-Proline/Lysine (T150PK) substitution within the P-domain forced DAXX delocalization, inducing replication stress and genome instability in a dominant-negative manner in our cell model. Gene expression analyses of the TCGA sarcoma dataset revealed that high SFPQ expression correlated with poor overall patient survival and reduced expression of the innate immunity gene signature. Conversely, low expression of the signature genes was associated with poor survival, which was further exacerbated in combination with high SFPQ expression. In conclusion, our data indicate that SFPQ

safeguards repeat stability by monitoring R-loop levels and recruiting DAXX-dependent histone H3.3 chaperone activity to maintain a stable chromatin template. In sarcoma patients, this pathway is predicted to suppress genome instability, limit the accumulation of cytoplasmic DNA species, and restrain activation of the cGAS–STING pathway, thereby fostering an immunosuppressive tumor microenvironment that promotes cancer aggressiveness. This scenario may significantly contribute to the limited applicability of immune checkpoint therapies in this tumor type. Future studies need to demonstrate whether targeting SFPQ-mediated R-loop suppression may promote innate immunity and increase the efficacy of immune pathway receptor and STING agonists in programming a tumor-suppressive microenvironment in sarcoma¹⁰².

Methods

Cell culture

Human cell lines used were obtained from ATCC and have not been cultured for longer than 2 months, minimizing the risk of cross-contamination or genetic drift. Based on the nature of the experiments and the short culture duration, authentication was not deemed necessary. Used cells were tested negative for mycoplasma contamination. Cells were cultivated at 37 °C, 5% CO₂. U-2 OS cells (osteosarcoma, ATCC HTB-96) were cultured in low glucose Dulbecco's modified Eagle's (DMEM) medium (Euroclone) with 10% fetal bovine serum (Corning), 1% L-glutamine (Gibco), 1% penicillin/streptomycin (Gibco). H1299 cells (carcinoma; non-small cell lung cancer, ATCC CRL-5803) were cultured in Roswell Park Memorial Institute (RPMI) medium (Euroclone) supplemented with 10% fetal bovine serum (Euroclone), 1% L-glutamine (Gibco), 1% penicillin/streptomycin (Gibco). MCF 10A (breast epithelial, ATCC CRL-10317) were cultured in high glucose Dulbecco's modified Eagle's (DMEM) medium and HAMS F12 (1:1 ratio, Euroclone), supplemented with 5% horse serum (Gibco), 1% L-glutamine (Gibco), 1% penicillin/streptomycin (Gibco), 10 μ g/ml insulin (Sigma), 0.5 μ g/ml hydrocortisone (Sigma), and 20 ng/ml EGF (Cell Guidance Systems). MCF-7 cells (ATCC HTB-22) were cultivated in Eagle's Minimum Essential Medium (EMEM) supplemented with 10% FBS (Gibco), 1% L-glutamine (Gibco), 1% penicillin/streptomycin (Gibco), and 10 μ g/ml insulin (Sigma). HEK293 (human embryonic kidney, ATCC CRL-1573) cells were cultured in high-glucose DMEM medium (Euroclone) supplemented with 10% fetal bovine serum (Euroclone), 1% L-glutamine (Gibco), 1% penicillin/streptomycin (Gibco). U-2 OS TetON-RNaseH1 cells, kindly provided by Professor Sébastien Britton (IPBS, Toulouse), were maintained in the presence of puromycin (0.25 μ g/ml; Sigma). Expression of mCherry-RNaseH1 was induced with doxycycline (2 μ g/ml; Sigma) for 24 h. MCF-10A and U-2 OS cells transduced with retroviral vectors (pBABE or pBABE-GFP-H2B) were selected with puromycin (1 μ g/ml). U-2 OS cells stably transfected with linearized control pcDNA or pCMVmyc-SFPQ deletion constructs were selected with puromycin (0.5 μ g/ml). Where indicated, U-2 OS cells were treated with H151 (2 μ M; Sigma) for 24 h, MG132 (20 μ M; Sigma) for 4 or 8 h, or hydroxyurea (2 mM; Sigma) for 24 h.

Plasmids

SFPQ vectors purchased from Addgene are listed in Supplementary data 9. SFPQ Δ P was generated by overlap extension PCR using primers containing KpnI and XhoI restriction sites for cloning the cDNA element into KpnI and XhoI sites of pCMV-Myc (ClonTech). pLV-eGFP was kindly provided by Prof. Giannino Del Sal (University of Trieste, Italy) and used for subsequent cloning. The derivative vector pAle was generated by inserting a fragment containing a nuclear localization sequence, a myc-epitope tag and multiple cloning site containing SpeI, NheI, XhoI sites using BsrGI. pAle allows the expression of fusion proteins containing a C terminal myc tag and N terminal eGFP SFPQ wild-type P domain or P domain containing T150delinsKP patient mutation (P dom nFS) were generated by PCR using primers

containing SpeI and XhoI sites for cloning the cDNA elements derived from FL SFPQ or SFPQ containing T150delinsKP patient mutation (SFPQ nFS) into SpeI-XhoI sites of pAle. HA-DAXX was generated by PCR using primers containing EcoRI and XbaI restriction sites for cloning the cDNA element into EcoRI and XbaI sites of pcDNA3.1-HA (Addgene). SFPQ mutants have been generated using the Quick-Change II XL Site-Directed Mutagenesis Kit (Agilent) following the manufacturing instructions. Briefly, forward and reverse primers are designed to be complementary one each other, harboring the desired mutation in the middle of the sequence, allowing for PCR amplification of the entire plasmid. His₆-myc SFPQ and GST-DAXX-HA were cloned using LIC cloning protocol, as previously described¹⁰³. Briefly, backbone vectors (pNIC-CTHF for His₆-Myc SFPQ, pGTVL1-SGC for GST-DAXX-HA) were enzymatically digested (pNIC-CTHF with BfuAI, pGTVL1-SGC with BsaI) and subsequently treated with T4 DNA polymerase to generate protruding ends. In parallel, PCR amplicons of desired inserts were treated with T4 DNA polymerase to generate protruding ends complementary to the one generated on the backbone. Backbone-insert were annealed at room temperature for 1 h, then used for transformation of competent *E. coli*. pNIC-CTHF was a gift from Opher Gileadi (Addgene plasmid # 26105). pGTVL1-SGC was a gift from Nicola Burgess-Brown (Addgene plasmid # 39188).

Transfection, transduction, and production of stable cell lines

Transient transfection: specific siRNAs (Supplementary Table 3) were transfected using RNAi-MAX Lipofectamine (Invitrogen) at 30 nM final concentration, according to the manufacturer's suggestions. Plasmids were transfected using Lipofectamine 2000 (Invitrogen), according to the manufacturer's suggestions.

Generation of stable cell lines: U-2 OS cells stably expressing myc-tagged SFPQ deletion mutants were produced by co-transfecting U-2 OS cells with I) a linearized plasmid containing SFPQ expression construct (1 µg) and II) a linearized pcDNA vector containing puromycin resistance (0.1 µg). 48 h post-transfection puromycin (0.5 µg/ml) was used to select for stable integrants.

Viral transduction: For lentiviral particle production, HEK 293 cells were co-transfected using Polyethylenimine (PEI, Sigma, 1 mg/ml) with transfer vectors and psPAX2 and pMD2-env packaging vectors. Virus-containing supernatant was collected 72 h post-transfection, centrifuged 5 min at 500 g, and filtered with a 0.45 µm low-protein-binding filter to remove cellular debris. The virus-containing medium was added to target cells with the addition of Polybrene (Sigma, 6 µg/ml). Puromycin (1 µg/ml) was used to select for stable integrants.

Western blotting

Whole-cell lysates were prepared as previously described and subjected to Western blotting according to standard procedures¹⁰⁴. Briefly, cells were lysed in RIPA buffer (50 mM Tris pH 7.4, 250 mM NaCl, 1% Triton-X, 1% DOC, 0.1% SDS) for 1 h at 4 °C. Subsequently, material was sonicated using a Fisherbrand Model 120 Sonic Dismembrator for 20 s at 20% amplitude and centrifuged for 10 min (14,000 g, 4 °C). Protein quantification was performed using a Bradford assay (Bio-Rad). Samples were mixed with 6x sample buffer (0.375 M Tris, pH 6.8; 12% SDS; 60% glycerol; 0.6 M DTT; 0.06% bromophenol blue) to a final 1x concentration and denatured at 95 °C for 5 min. Proteins were resolved on homemade SDS-PAGE gels in Tris-Glycine running buffer (25 mM Tris, 190 mM glycine, 0.1% SDS) and transferred to nitrocellulose membranes (Cytiva) using transfer buffer (25 mM Tris, 190 mM glycine, 0.01% SDS, 20% methanol). Membranes were blocked with blocking buffer (5% BSA in 1xPBS containing 0.1% Tween-20) and incubated overnight at 4 °C with primary antibodies diluted in blocking buffer. The following day, membranes were washed three times with 1xPBS containing 0.1% Tween-20, incubated with HRP-conjugated secondary antibodies diluted in blocking buffer, and washed three times again with 1xPBS containing 0.1% Tween-20. Signal

detection was performed using Clarity Western ECL substrate (Bio-Rad) and imaged with an iBright FL1500 Imaging System (Invitrogen). Primary and secondary antibodies are listed in Supplementary data 10.

Protein immunoprecipitation

Cell extracts were prepared using IP buffer (50 mM Tris, pH 8.0; 150 mM NaCl; 1% NP-40; 5 mM EDTA; 5% glycerol) supplemented with 1 mM PMSF (Sigma), 1x protease inhibitor cocktail (PIC; Sigma), and 1 mM NaF (Sigma). Extracts were passed several times through a small syringe to ensure efficient nuclear disruption. For each immunoprecipitation, 800 µg of whole-cell extract was incubated with specific antibodies (listed in Supplementary data 10) overnight at 4 °C. Antibody-antigen complexes were captured using 25 µl Protein A Dynabead slurry (Invitrogen) by rocking samples for 2 h at 4 °C. Beads were washed four times with IP buffer, and bound proteins were eluted in 2x sample buffer (125 mM Tris, pH 6.8; 0.5% SDS; 10% glycerol; 5% 2-mercaptoethanol). Eluted proteins were subsequently analyzed by Western blotting.

Mass spectrometry

Cell lysates were prepared with ice-cold MS Lysis Buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 5 mM EDTA, 1 mM DTT, 5% glycerol, 1% NP-40, supplemented with 1 mM PMSF, 1 mM beta-glycerophosphate, 1 mM Na₃VO₄, 5 mM NaF, and 1x PIC). After centrifugation and preclearing, lysates were incubated at 4 °C with specific antibodies or rabbit IgG as negative control (Supplementary data 10). After 2 h, protein-A agarose beads (Santa Cruz) were added and incubated overnight. The resin was then washed in MS Lysis Buffer four times and bound proteins were eluted in 2x sample buffer. Immunoprecipitated proteins were separated by SDS-PAGE and stained using mass spectrometry-compatible silver staining (SilverQuest™ Kit, Invitrogen) or Coomassie Brilliant blue R-250 (42660, Sigma). Selected lanes were excised from gels, destained and washed twice in 50% ACN with 50 mM ammonium bicarbonate, and dehydrated in 100% ACN. In-gel reduction was performed by incubating lanes in a 10 mM DTT, 100 mM ammonium bicarbonate solution for 30 min at 56 °C and alkylation in 55 mM iodoacetamide, 100 mM ammonium bicarbonate for 20 min at room temperature. Protein samples were dehydrated again in 100% ACN and digested by rehydrating gel pieces in 50 mM ammonium bicarbonate containing 4 ng/µl of trypsin (#V5111, Promega) overnight at 37 °C. Tryptic peptides were desalted and concentrated using ZipTip mC18 pipet tips (Millipore) and co-eluted onto the MALDI target in 1 µl of α-cyano-4-hydroxycinnamic acid matrix (5 mg/ml in 50% ACN, 0.1% TFA). Mass spectra were acquired over a mass range of 800–4000 m/z (Nd:YAG laser at 355 nm, 40 Shots/Sub-Spectrum for 2000 Total Shots/Spectrum) by reflectron positive mode on an Applied Biosystems 4800 Proteomics Analyzer mass spectrometer (Applied Biosystems) and calibrated using a standard mixture (Mass Standards Kit, AB SCIEX). MS/MS spectra were acquired in positive mode (Nd:YAG laser at 355 nm, 40 Shots/Sub-Spectrum for 4000 Total Shots/Spectrum) and MS/MS calibration was achieved by using the default calibration method. Protein identifications were performed with the ProteinPilot™ software (version 2.0.1; Applied Biosystems) using the Paragon™ algorithm as the search engine. Each MS/MS spectrum was searched against the Uniprot/SwissProt database. The search parameters allowed for cysteine modification by iodoacetamide and biological modifications programmed in the algorithm (i.e., phosphorylations, semitryptic fragments, etc.). The detected protein threshold (ProtScore) in the software was set to 1.3 to achieve 95% confidence interval.

Immunofluorescence

Cells were fixed in 4% paraformaldehyde (PFA) in 1xPBS for 15 min, followed by treatment with citrate buffer (0.1% w/v; 0.5% Triton X-100) for 5 min at room temperature. Subsequently, cells were then blocked

for 1 h in blocking solution (3% BSA, 0.1% Tween-20 in 1x PBS) and incubated with primary antibodies (listed in Supplementary data 10) diluted in blocking solution for 2 h at room temperature. After incubation, cells were washed three times for 5 min each with washing solution (0.5% BSA, 0.1% Tween-20 in 1xPBS) and subsequently incubated with secondary antibodies (Supplementary data 10) diluted in washing solution for 1 h at room temperature. Slides were washed twice for 10 min in washing solution, stained with DAPI (1 µg/ml; Sigma) in 1xPBS containing 0.1% Tween-20, and mounted with ProLong (Invitrogen). For immunofluorescence analysis using monoclonal S9.6 antibodies, cells were fixed and permeabilized with ice-cold methanol for 10 min followed by acetone for 1 min on ice, as previously described¹⁰⁵. Blocking, antibody dilution and washing solutions were prepared as aforementioned, using 4xSSC instead of 1xPBS. To evaluate target specificity of the S9.6 antibody, U-2 OS cells fixed on glass slides were treated with RNaseT1 (NEB, 1U/µl; NEBuffer 2), RNaseT3 (NEB, 1U/µl; NEBuffer 2) or RNaseH1 (NEB, 1U/µl; NEBuffer 2) at 37 °C for 30 min. Samples were incubated at 65 °C for 20 min to inactivate the enzymes, washed in 1xPBS and processed following the immunofluorescence protocol. For immunofluorescent detection of cytosolic dsDNA with dsDNA-specific antibodies, cells were fixed in 4% PFA, permeabilized with 0.1% saponin (Sigma) in 1xPBS on ice for 5 min and subjected to the standard immunofluorescence protocol. For conventional immunofluorescence analysis, a Leica DM4000B microscope was used; quantitative analysis of immunofluorescence was performed with ImageJ (1.54i). A Student's *t* test was used to calculate statistical significance.

A ZEISS ELYRA 7 Structured Illumination Super-Resolution Microscope with ZEN Black software was used for high-resolution microscopy.

Immuno-DNA FISH

Cells were washed in 1xPBS and fixed in 4% paraformaldehyde (PFA) in 1xPBS for 20 min at room temperature, followed by permeabilization with 1xPBS containing 0.1% Triton X-100 for 7 min at room temperature. Cells were then blocked with 5% BSA in 1xPBS at 37 °C for 20 min, and immunofluorescence staining was performed using specific antibodies (listed in Supplementary data 10). Antibody-antigen complexes were post-fixed at room temperature for 2 min with 4% PFA in 1xPBS, dehydrated through an ethanol series (70%, 90%, 100%; 5 min each), and air-dried. DNA-FISH was carried out in hybridization solution [10 mM Tris-HCl (pH 7.0) containing 70% (v/v)] deionized formamide, 0.25% (w/v) blocking reagent (Sigma), 5% (v/v) of MgCl₂ buffer solution (25 mM MgCl₂, 9 mM citric acid, 82 mM Na₂HPO₄), and 0.25 µg/ml FISH probe]. Samples were by denatured at 80 °C for 3 min. Subsequently, slides were incubated for 2 h at room temperature in a light-protected, humid chamber. Following hybridization, slides were washed in FISH washing solution [50% formamide, 10 mM Tris-HCl (pH 7.2), 0.1% BSA], stained with DAPI (1 µg/ml; Sigma) in 1xPBS containing 0.1% Tween-20, dehydrated again through an ethanol series (70%, 90%, 100%; 5 min each), air-dried, and mounted with ProLong (Invitrogen). C₀t-1 FISH probes were generated using Human C₀t-1 DNA (Invitrogen) using the FISH Tag DNA kit (Invitrogen), following manufacturer's instructions. 5'-TYE™563 labeled pericentromeric SATIID probes (GATCGAATGGAATCTGAATGGAA) were purchased from Panagene. Telomere probes (TelG-Cy5) were purchased by Panagene. A ZEISS ELYRA 7 Structured Illumination Super-Resolution Microscope with ZEN Black software was used for high-resolution microscopy.

Chromatin immunoprecipitation

Chromatin immunoprecipitation was performed as previously described¹⁰⁶. Briefly, 1 × 10⁷ cells per immunoprecipitation were fixed in 1% formaldehyde (Sigma) in 1xPBS for 15 min at room temperature. The reaction was quenched with 140 mM glycine (final concentration) for 5 min at room temperature. Cells were washed twice with ice-cold

1xPBS and scraped into 1xPBS supplemented with 1× protease inhibitor cocktail (Sigma-Aldrich), 1 mM NaF, and 1 mM PMSF. Cell pellets were collected by centrifugation at 4000 *g* for 5 min at 4 °C. Cells were lysed in Lysis Buffer I (50 mM HEPES, pH 7.5; 10 mM NaCl; 1 mM EDTA; 10% glycerol; 0.5% NP-40; 0.25% Triton X-100) by rocking at 4 °C for 10 min, followed by centrifugation at 1500 *g* for 5 min at 4 °C. The pellet was resuspended in Lysis Buffer II (10 mM Tris-HCl, pH 8.0; 200 mM NaCl; 1 mM EDTA; 0.5 mM EGTA) and centrifuged at 1000 *g* for 10 min at 4 °C. The pellet was then resuspended in Lysis Buffer III (10 mM Tris-HCl, pH 8.0; 200 mM NaCl; 1 mM EDTA; 0.5 mM EGTA; 0.1% sodium deoxycholate; 0.5% N-lauroylsarcosine). Chromatin was sonicated using a BioRuptor (Diagenode) to obtain DNA fragments of 150–300 bp. Chromatin was diluted in Equilibration Buffer (10 mM Tris-HCl, pH 8.0; 100 mM NaCl; 1 mM EDTA; 1.66% Triton X-100; 0.166% sodium deoxycholate) at a 1:1.5 ratio (1 ml chromatin with 1.5 ml buffer) to a final volume of ~900 µl. An aliquot of processed chromatin was reverse-crosslinked (see below) and amount of recovered DNA was determined. A quantity of chromatin corresponding to 1 µg purified DNA was used in downstream immunoprecipitation experiments. Specific antibodies or rabbit/mouse control IgGs (Supplementary data 10) were added, and samples were incubated overnight at 4 °C with rocking. Protein A Dynabeads (Invitrogen) were blocked overnight in 1xPBS containing 0.5% BSA and 5 mM EDTA. The following day, antibody-antigen complexes were captured using 25 µl Protein A Dynabeads per sample, rocking for 2 h at 4 °C. Beads were sequentially washed in Low Salt Wash Buffer (20 mM Tris-HCl, pH 8.0; 150 mM NaCl; 2 mM EDTA; 0.1% SDS; 1% Triton X-100; 5 min), High Salt Wash Buffer (20 mM Tris-HCl, pH 8.0; 500 mM NaCl; 2 mM EDTA; 0.1% SDS; 1% Triton X-100; 5 min), and twice in TE buffer (10 mM Tris-HCl, pH 8.0; 1 mM EDTA; 5 min each). Proteins were eluted twice in Elution Buffer (10 mM Tris-HCl, pH 8.0; 300 mM NaCl; 5 mM EDTA; 0.5% SDS) at 37 °C for 15 min each, shaking at 850 rpm. Immunoprecipitated chromatin was de-crosslinked by adding 200 mM NaCl and incubating at 65 °C overnight. Samples were treated with RNase A (20 µg; Invitrogen) and Proteinase K (40 µg; Invitrogen) for 1 h at 45 °C, followed by phenol-chloroform extraction and ethanol precipitation. DNA was resuspended in TE buffer. Quantitative real-time PCR was performed using the CFX Connect Real-Time PCR Detection System (Bio-Rad) and iTaq Universal SYBR Green Supermix (Bio-Rad), according to the manufacturer's instructions. The PCR protocol for the amplification of telomere repeats has been described in ref. 107. For data analysis, samples were normalized on input, IgG sample was subtracted as noise, siControl was set as 1. For ChIP-seq experiment, DNA sample were purified using QIAquick PCR Purification Kit (Qiagen). Purified DNA (10 ng) was used for the preparation of a TruSeq ChIP Library (Illumina). 2x150pb paired-end sequencing was performed on an Illumina HiSeqX platform; data throughput: 18 Gb. Library preparation and sequencing service was performed by MacroGen Europe B.V.

DNA:RNA hybrid immunoprecipitation (DRIP)

1 × 10⁷ cells were scraped into 1xPBS, centrifuged for 5 min at 4000 *g*, 4 °C, and lysed in Lysis Buffer (1% SDS, 20 mM Tris-HCl pH 7.5, 40 mM EDTA pH 8.0, 100 mM NaCl, ddH₂O). TE buffer (100 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0) was added at a 1:1 ratio, and samples were supplemented with Proteinase K (Invitrogen) to a final concentration of 150 µg/ml. Samples were incubated at 37 °C overnight. Genomic DNA was purified by phenol-chloroform extraction, ethanol precipitation, and resuspended in TE buffer. Chromatin was sonicated using a BioRuptor (Diagenode) to obtain DNA fragments ranging from 500–1200 bp. For each sample, 4 µg of genomic DNA was either treated with recombinant RNase H1 (New England Biolabs; 3 units per 1 µg DNA at 37 °C) or left untreated. Protein A Dynabeads (Invitrogen) were blocked overnight in 1xPBS containing 0.5% BSA and 5 mM EDTA. The following day, antibody-bead conjugation was performed by incubating 2 µg of S9.6 monoclonal antibody (Kerafast) with 25 µl Protein A

Dynabeads per sample. Samples were rocked for 4 h at 4 °C in IP Buffer (50 mM HEPES/KOH pH 7.5, 140 mM NaCl, 5 mM EDTA pH 8.0, 1% Triton X-100, 0.1% sodium deoxycholate). Genomic DNA was then added to antibody-coupled beads and incubated overnight at 4 °C with rocking. The following day, samples were washed (5 min each) sequentially in IP buffer, Wash Buffer I (50 mM HEPES/KOH pH 7.5, 500 mM NaCl, 5 mM EDTA pH 8.0, 1% Triton X-100, 0.1% sodium deoxycholate), Wash Buffer II (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, 1 mM EDTA pH 8.0), and twice in TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). DNA was eluted twice in Elution Buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0, 1% SDS) at 65 °C for 15 min each, shaking at 850 rpm. DNA was purified by phenol–chloroform extraction, ethanol precipitation, and resuspended in TE buffer. Enrichment of RNA:DNA hybrids at target regions was validated by quantitative real-time PCR using the CFX Connect Real-Time PCR Detection System (Bio-Rad) and iTaq Universal SYBR Green Supermix (Bio-Rad), according to the manufacturer's instructions. The PCR protocol for telomere repeat amplification has been previously described¹⁰⁷. For data analysis, samples were normalized to input DNA; RNase H1-treated DNA was used to confirm immunoprecipitation specificity, and siControl RNase H1-untreated samples were set to "1."

Gene expression analysis by quantitative RT-PCR

Cells were washed with 1xPBS, and total RNA was prepared using Trifast (Euroclone) following manufacturer's instructions. Reverse transcription with random primers was performed by using QuantiTect® Reverse Transcription Kit (Qiagen) following manufacturer's instructions. Quantitative PCR (RT-qPCR) was performed using gene specific primers (Supplementary data 8) and the iTaq™ Universal SYBR® Green Supermix (BioRad) in a CFX Connect Real-Time PCR Detection System (BioRad). Specificity of PCR products was routinely checked by melting curve analysis and agarose gel electrophoresis. The PCR protocol for the amplification of TERRA cDNA has been previously described³.

RNA sequencing (RNA-Seq)

Total RNA was prepared using the Norgen Biotech RNA extraction Kit following manufacturer's instructions. A TruSeq stranded mRNA library (Illumina) was sequenced on a NovaSeq 6000 RNA Seq platform. 2 × 150 bp paired end sequencing with 18 Gb throughput was performed by Macrogen Europe B.V.

Purification of recombinant proteins

E. coli Rosetta 2 (DE3) cells (Novagen) were transformed with expression vectors for human SFPQ and DAXX (pNIC-CTHF His₆-myc SFPQ, pGTVL1-SGC-GST-DAXX-HA). Cells were grown to an OD₆₀₀ of 0.6–0.8 in media containing appropriate antibiotics (chloramphenicol, 34 µg/ml for both plasmids; ampicillin, 100 µg/ml for pNIC-CTHF His₆-myc SFPQ; kanamycin, 50 µg/ml for pGTVL1-SGC GST-DAXX-HA) and subsequently transferred to 18 °C. Protein expression was induced with 0.1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG; Sigma). Sixteen hours post-induction, cells were harvested by centrifugation at 3000 g for 15 min. To purify recombinant human SFPQ, cell pellets were resuspended in Lysis Buffer I (50 mM Tris-HCl, pH 7.5; 300 mM NaCl; 5% glycerol; 5 mM imidazole; 1 mM TCEP) supplemented with DNase I (0.1 mg/ml; Sigma), PMSF (1 mM; Sigma), benzamide (15 µg/ml; Sigma), leupeptin (4 µg/ml; Sigma), and aprotinin (2 µg/ml; Sigma). Cells were disrupted by sonication, and crude extracts were centrifuged for 1 h at 30,000 g. The soluble fraction was incubated with 1 ml Ni-NTA resin (Qiagen) for 1 h at 4 °C. Resin was washed with 20 volumes of lysis buffer, and bound protein was eluted with lysis buffer containing 300 mM imidazole. Appropriate fractions were pooled and concentrated using Amicon Ultra-15 centrifugal filter units (Merck). The protein was further purified on a Superdex 200 10/300 GL column (Cytiva) equilibrated in SEC Buffer (20 mM Tris-HCl, pH 8.0; 300 mM

NaCl; 0.5 mM TCEP; 10% glycerol). Purified protein was flash-frozen in liquid nitrogen and stored at –80 °C. To purify recombinant human DAXX protein, cell pellets were resuspended in Lysis Buffer II (50 mM NaH₂PO₄; 300 mM NaCl; 10 mM imidazole; 1 mM TCEP) supplemented with DNase I (0.1 mg/ml), PMSF (1 mM), and complete EDTA-free protease inhibitor tablets (Roche). Cells were disrupted by sonication, and crude extracts were centrifuged at 30,000 g for 1 h. The soluble fraction was incubated with 1 ml Glutathione Sepharose 4B resin (Cytiva) for 1.5 h under rotation at 4 °C. Beads were sequentially washed with lysis buffer II supplemented with decreasing concentrations of KCl (1.0, 0.5, 0.25, and 0 M). Bound protein was eluted with lysis buffer II containing 25 mM glutathione. Appropriate fractions were pooled and concentrated using Amicon Ultra-15 centrifugal filter units (Merck). For further purification, recombinant DAXX was applied to a Superdex 200 10/300 GL column (Cytiva) equilibrated in SEC Buffer. Purified protein was flash-frozen in liquid nitrogen and stored at –80 °C.

Protein pull-down assays

To validate SFPQ–DAXX interaction, pull-down experiments were carried out using recombinant His₆-myc-SFPQ and GST-DAXX-HA. 10 µg of purified His₆-myc-tagged SFPQ were added to 10 µl c-myc magnetic beads (Thermo Scientific) in SEC buffer (20 mM Tris-HCl, pH 8.0; 300 mM NaCl; 0.5 mM TCEP; 10% glycerol) at 4 °C under rotation for 2 h. Beads were washed three times with SEC Buffer to eliminate excess SFPQ protein. Recombinant GST-DAXX-HA (5 µg) was incubated with immobilized recombinant SFPQ at 4 °C under rotation for 2 h in SEC buffer. Beads were then washed four times with SEC Buffer. Alternatively, 10 µg recombinant GST-DAXX-HA was immobilized to Glutathione Sepharose 4B beads (10 µl) under rotation for 1 h at 4 °C in SEC buffer. Excess recombinant DAXX protein was removed by three washes in SEC buffer. 5 µg of recombinant His₆-myc-tagged SFPQ was incubated with immobilized recombinant GST-DAXX-HA at 4 °C under rotation for 2 h in SEC buffer. Beads were washed four times with SEC Buffer. Beads were resuspended in 2x Sample buffer (2% SDS, 10% glycerol, 0.004% bromophenol blue, 0.125 M Tris-Cl, pH 6.8, 5% DTT) and boiled for 10 min at 95 °C. Samples were analyzed by western blot using an anti-myc (SFPQ) and anti-HA(DAXX) antibodies (Supplementary data 10).

Preparation of probes for EMSA

RNA and DNA oligonucleotides were chemically synthesized and purified by reverse-phase high pressure liquid chromatography and PAGE (Biomers.net). Selected oligonucleotides were 5' labeled with fluorescent 6-Carboxyfluorescein (6-FAM). Used DNA and RNA oligos are listed in Supplementary data 1. Oligonucleotides were resuspended in TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8.0) at a final concentration of 100 µM and stored at –20 °C. For fork-DNA probes (DIL:D3L) the fluorescent strand and the complementary strand were annealed at a 1:3 molar ratio in 20 mM Tris pH 8.0 and 100 mM NaCl. 100 µl reactions including 10 µM of each strand were heated to 95 °C for 5 min and gradually cooled to 15 °C at a rate of 1 °C min⁻¹ using a PCR thermal cycler (Eppendorf). For *D-loop* (D4:D9:D11) and *R-loop* (R4:D9:D11) the molar ratio used were 1:1.25:2.5 in 6 mM Tris-HCl pH 7.5, 7 mM MgCl₂, 50 mM NaCl and 1 mM DTT. The annealing method used for D-loop and R-loop was heating at 99 °C for 5 min followed by incubations at 67 °C for 1 h, at 37 °C for 30 min and at 25 °C 3–4 h or overnight¹⁰⁸. Probes were run on native 10% acrylamide (29:1), 1xTBE gels at room temperature in 1x TBE running buffer. Nucleic acid substrates were eluted from gel slices by dialysis in 1xTE buffer.

Electrophoretic mobility shift assays (EMSA)

Nucleic acid substrates (10 nM) were incubated with increasing amounts of recombinant His₆-myc-SFPQ (0, 200, 400, 800, 1600 nM). Binding was performed in EMSA buffer (20 mM Tris-HCl, pH 7.5, 100 mM KCl, 0.5 mM MgCl₂, 0.5 mM EDTA, 0.2 mM DTT, 1% glycerol)

at room temperature for 30 minutes. RNase inhibitor (RNaseOUT, Thermo Fisher) was added to reactions containing RNA oligonucleotides at a final concentration of 0.16 units/ μ l. When indicated, competitor nucleic acids were added. Reactions were run on a non-denaturing 5% polyacrylamide gel at room temperature in 1xTBE buffer for 40 min. Fluorescent-labeled substrates were detected using a fluorescent scanner (ImageQuant, GE Healthcare). Quantification of protein-bound nucleic acid was performed using ImageJ software.

Sister chromatid exchange assays

Experimental cells were grown in the presence of 5'-bromo-2'-deoxyuridine (BrdU, Sigma, 10 μ M) for two rounds of DNA replication. Colcemid was added during the final period of BrdU treatment (0.2 μ g/ml, 4 h). Single cells were recovered by trypsinization and incubated with hypotonic solution (46.5 mM KCl, 8.5 mM Na Citrate) at 37 °C for 30 min. Cells in suspension were fixed by slowly adding MeOH/Acetic Acid (ratio 3:1) under soft mixing. Cells were centrifuged (1000 *g*, 5 min, 4 °C) and fixation procedure was repeated twice. Finally, cells were dropped on glass slides to obtain metaphase spreads. Differential staining of sister chromatids was performed, as previously described¹⁰⁹. Briefly, metaphase spreads were stained with Hoechst 33,258 (10 μ g/ml in 1xPBS, 20 min), washed in Sorensen buffer (0.1 M Na₂HPO₄, 0.1 M NaH₂PO₄, pH 6.8) and exposed to UV treatment (UV transilluminator, BioRad) for 30 min to degrade preferentially DNA with incorporated BrdU, thus allowing a differential staining of sister chromatids after Giemsa staining. Metaphase spreads were incubated at 50 °C for 1 h in 1x SSC solution and subsequently dyed in Sorensen buffer supplemented with 10% Giemsa stain solution (Sigma, 0.4%, w/v). Slides were washed in 1xSSC and mounted in ProLong (Invitrogen). Chromosomal aberrations, such as condensation and cohesion defects, recombination and breaks were quantified by visual inspection.

Immunofluorescence on native metaphase chromosome spreads

For chromosome spreads, asynchronous cells were treated with 0.2 μ g/ml Colcemid for 4 h. Cells were collected by mitotic shake-off, as previously described⁶³. Mitotic cells were resuspended in 65 mM KCl buffer and allowed to swell for 10 min at room temperature. Subsequently, cells were spun onto glass slides using a Cytopro cytospin-centrifuge (Wescor) at 250 *g* for 5 min. Slides were incubated in KCM buffer (10 mM KCl, 20 mM NaCl, 10 mM Tris-HCl, pH 8, 0.5 mM EDTA, 0.1% Triton X-100) for 10 min. Primary antibodies (Supplementary data 10) were diluted in 1% BSA/KCM buffer. Slides were incubated with primary antibodies in a wet chamber at room temperature for 2 h. Slides were washed three times with KCM buffer for 5 min each, and incubated with secondary antibodies (Supplementary data 10) in 1% BSA/KCM buffer in a wet chamber at room temperature for 1 h. Cells were subsequently washed three times with KCM buffer, 5 min each, fixed with 4% paraformaldehyde (PFA, in 1xPBS) at room temperature for 15 min. Slides were mounted in ProLong (Invitrogen). All buffers used were supplemented with NaF (Sigma, 1 mM) and sodium orthovanadate (Na₂VO₄, Sigma, 1 mM) to prevent dephosphorylation of proteins. Images were captured using a classic immunofluorescence microscope (Leica DM4000B) and were analyzed by visual inspection.

RNA-fluorescence in situ hybridization (RNA-FISH)

Cells were permeabilized at room temperature by incubation with Cytobuffer (100 mM NaCl, 300 mM Sucrose, 3 mM MgCl₂, 10 mM Pipes pH 6.8) for 30 s, followed by treatment with Cytobuffer supplemented with 0.5% Triton-X for 30 s and a final wash in Cytobuffer for 30 s. Subsequently, cells were fixed for 10 min at room temperature in ice-cold 4% paraformaldehyde (in 1xPBS). Fixed cells were dehydrated by two washes in ice-cold 70% ethanol for 2 min, one wash in ice-cold 90% ethanol for 1 min, and one wash in ice-cold 100% ethanol

for 1 min. Slides were allowed to air-dry and were incubated overnight with an AlexaFluor555 labeled human C₀t-1 probe in RNA FISH Hybridization Buffer (2x SSC, 50% formamide) in a humid chamber (2x SSC, 50% formamide) at 37 °C. Subsequently, slides were washed once in 2x SSC for 3 min at room temperature and three times in 2x SSC for 5 min at 37 °C. Slides were transferred to 4x SSC buffer containing DAPI (Sigma, 1 μ g/ml). Slides were mounted in ProLong (Invitrogen). Images were captured using an immunofluorescence microscope (Leica DM4000B). For spot signals intensity of interphase nuclei, TFL-TELO software was used. For high-resolution microscopy, a ZEISS ELYRA 7 Structured Illumination Super-Resolution Microscope was used.

Combined immuno-RNA FISH on interphase cells

Cells were permeabilized with Cytobuffer (100 mM NaCl, 300 mM Sucrose, 3 mM MgCl₂, 10 mM Pipes, pH 6.8) and Cytobuffer supplemented with Triton X-100, as described for RNA-FISH and subsequently subjected to the immunofluorescence protocol. Primary and secondary antibodies are listed in Supplementary data 10. After washing, antibody-antigen conjugates were fixed in 4% paraformaldehyde (in 1xPBS) at room temperature for 2 min, washed in 2x SSC and dehydrated in ethanol series (ice-cold 70%, 90%, 100%; 5 min each) followed by RNA-FISH. A ZEISS ELYRA 7 Structured Illumination Super-Resolution Microscope was used to detect co-localization events.

Combined immuno-RNA FISH on native metaphase chromosomes

Immuno-RNA FISH on native chromosome spreads was carried out with modification to Johnson et al.¹¹⁰. Briefly, asynchronous cells were treated with 0.2 μ g/ml colcemid for 4 h, collected by trypsinization, resuspended in hypotonic solution (75 mM KCl), and incubated at room temperature (RT) for 20 min. Cells were spun onto glass slides pre-treated with poly-L-lysine (50 μ g/ml; Sigma). Cells were permeabilized in KCM buffer (10 mM Tris-HCl, pH 8.0; 120 mM KCl; 20 mM NaCl; 0.5 mM EDTA; 0.1% Triton X-100) for 5 min at RT, then fixed in 4% paraformaldehyde (PFA in 1xPBS) for 10 min. Slides were blocked in KCM buffer supplemented with 2% BSA for 1 h at RT, followed by overnight incubation at 4 °C with primary antibodies (Supplementary data 10) diluted in blocking buffer. The next day, slides were washed three times in KCM buffer and incubated with secondary antibodies (Supplementary data 10) diluted in KCM buffer with 2% BSA for 1 h at RT. Slides were washed three times in KCM buffer, fixed again in 4% PFA (1xPBS) for 10 min, and permeabilized in CSK buffer (100 mM NaCl; 300 mM sucrose; 3 mM MgCl₂; 10 mM PIPES, pH 6.8) supplemented with 0.5% Triton X-100 on ice for 10 min. Slides were dehydrated in an ethanol series (70%, 90%, 100%; 2 min each, on ice), air-dried for 10 min at RT, and incubated with RNA-FISH probes (0.1 μ M) in hybridization mix (50% formamide, 2x SSC) overnight at 37 °C in a humid chamber. Slides were washed three times in Washing Buffer I (50% formamide, 2x SSC, 0.05% Tween-20) for 5 min at 37 °C, followed by three washes in Washing Buffer II (2x SSC, 0.05% Tween-20) for 5 min at 37 °C. Chromosomes were counterstained with DAPI (1 μ g/ml; Sigma) in Washing Buffer II. Excess DAPI was removed with Washing Buffer III (4x SSC, 0.05% Tween-20), and slides were mounted with ProLong (Invitrogen). Images were captured using a classic immunofluorescence microscope (Leica DM4000B) and were analyzed by visual inspection.

RNA:DNA hybrid dot-blot analysis

1×10^7 cells were scraped into 1xPBS, centrifuged for 5 min at 4000 *g*, 4 °C, and lysed in Lysis Buffer (1% SDS, 20 mM Tris-HCl pH 7.5, 40 mM EDTA pH 8.0, 100 mM NaCl). Subsequently, TE buffer (100 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0) was added at a 1:1 ratio. Proteinase K (Invitrogen) was added to a final concentration of 150 μ g/ml, and samples were incubated overnight at 37 °C with shaking. Genomic DNA

was purified by phenol–chloroform extraction, ethanol precipitation, and resuspended in TE buffer. DNA was fragmented by sonication using a BioRuptor (Diagenode) to obtain fragments of 150–300 bp. For each sample, 4 µg of genomic DNA was either treated with recombinant RNase H1 (New England Biolabs; 3 units per 1 µg DNA, 37 °C overnight) or left untreated. Samples were diluted in 2x SSC, spotted onto nylon membranes (PerkinElmer), and UV-crosslinked (UVP/Analytik Jena CX-2000; two rounds at 1,200,000 µJ/cm²). Membranes were blocked in Blocking Buffer (5% BSA in 1xPBS) and subjected to standard Western blotting using the S9.6 antibody to detect RNA:DNA hybrids. Total genomic DNA was visualized with SybrGold (Invitrogen).

Microscopy

Cells subjected to immunofluorescence, RNA-FISH, DNA-FISH, or immunofluorescence combined with RNA-FISH were analyzed using a Leica DM4000B microscope equipped with a Leica DFC420C digital camera. Images were captured using Leica Application Suite (LAS) imaging software. For live-cell imaging, confocal analysis was performed using a Nikon Eclipse C1si confocal microscope system. Images were processed by using ImageJ 1.46r (NIH, Bethesda, USA). In all experiments, single focal planes were used for analysis. A ZEISS ELYRA 7 Structured Illumination Super-Resolution Microscope with ZEN Black software was used for super-resolution microscopy. Colocalization analysis was performed in Fiji (ImageJ) using the JaCoP plugin. For each field, at least three DAPI-defined nuclear ROIs were analyzed. Manders' M2 (fraction of signals overlapping) was computed with Costes' automatic thresholding on a per-nucleus basis and averaged per image to yield one M2 value per field. $n = 10$ fields per condition were used for statistics. Where indicated, Costes' randomization (≥ 100 permutations) and Van Steensel's cross-correlation analysis were applied to assess non-random overlap and spatial correlation.

Live-cell imaging analysis

Cell cycle progression of H2B-GFP expressing cells was followed for 14–16 h using Live imaging microscope. Pictures of three separate optical fields were acquired every 10 min. Defects in mitotic progression were quantified by visual inspection of three independent experiments.

Cell cycle analysis

Cell cycle distribution was assessed by propidium iodide (PI) staining followed by flow cytometry. Briefly, 1×10^6 cells were harvested, washed with 1xPBS, fixed with cold 70% ethanol, and incubated at 4 °C for at least overnight. Fixed cells were centrifuged at 250 g for 5 min to remove residual ethanol, washed twice with 1xPBS, and rehydrated in 1xPBS at 4 °C for 1 h. Cells were subsequently resuspended in PI staining solution (10 µg/ml PI in 1xPBS; Sigma-Aldrich) supplemented with RNase A (100 µg/ml; Thermo Fisher Scientific) and incubated at 4 °C overnight. Flow cytometry assays were performed using an Attune NxT® flow cytometer (Thermo Fisher Scientific) with acoustic focusing technology, equipped with the standard optical bench configuration, with one blue laser (488 nm). PI fluorescence was collected at BL3 (695/40) PMT. Hierarchical gates and an acquisition flow rate of 100 µl/min were set to avoid coincident events. After recording at least 20,000 events per run, data were saved as FCS files and analyzed using FCS Express V7 De Novo Software with the “Multicycle” function for cell cycle analysis.

Bioinformatics analysis—ChIP-Seq

ChIP-seq analysis was performed using the ChIP-AP pipeline (ref. 111), performing quality control (FastQC), read trimming and filtering (Bbduk and Trimmomatic), genome mapping (BWA-MEM), peak calling (Genrich, HOMER, MACS2, SICER2), peak merging and annotation (HOMER). Using R/Bioconductor (ref. 112) environment, peaks that were called by three out of four peak calling methods were considered

as present. Peaks were considered differentially expressed when having p -value < 0.05 for at least three methods. In particular, for HOMER, MACS2, and SICER2, the Fold Change must be greater than 1 or for Genrich the enrichment score must be positive (upregulated peaks). Overlaps between peaks were calculated with findOverlaps function (IRanges package, ref. 113) with a minimum overlap of 100 bp. Distribution of genomic region annotation was visualized with piecharts generated with the PieDonut function (webr package). Boxplots of the mean of Log₂Fold Change were performed with ggboxplot function (ggpubr package), p -value was calculated with geom_pwc (ggpubr package) function with standard Wilcoxon test, p -adjusted calculation was performed using “fdr” flag. The number of peaks belonging to each genomic region was calculated using the stat_n_text function (EnvStats package). Venn of overlapping peaks was performed using the makeVennDiagram function (ChIPpeakAnno package) using a minimum overlap of 100 bp. For area under curve (AUC) analysis, peak calling was followed by quantification of enrichment signals using the area under the curve (AUC) for each peak. The AUC was calculated as the integral of the normalized ChIP-seq signal (per-base bigWig coverage values) across the full width of each peak. The log₂-transformed average enrichment signal (AUC + 1) was used for downstream comparisons. Boxplots were generated to visualize signal distribution across experimental conditions.

Bioinformatics analysis—RNA-Seq

RNA sequencing was performed in triplicate for individual conditions. Quality of raw sequence files was checked via FastQC¹¹⁴. Transcript quantification was conducted with STAR (version v.2.7.9a; ref. 115) using Ensembl GRCh38 GTF file and genome version (accessed on September 2022). The generated gene counts were consequently analyzed using the R package DESeq2¹¹⁶. The normalized count matrix was obtained from variance stabilizing transformation (VST) method as implemented in DESeq2 package. In order to explore high-dimensional data property, among the available algorithms, Principal Component Analysis (PCA) coupled with a dimensionality reduction algorithm was used to scale data and visualized with ggplot2¹¹⁷. In order to select only the statistically significant changing genes between comparisons of interest, a differential gene expression analysis was performed. The differentially expressed genes (DEGs) were selected with a p -adjusted cut off of 0.05 and a log₂ Fold Change (LFC) value greater than 1.5 (up-regulated DEGs) or lower than -1.5 (down-regulated DEGs). P -value was adjusted for multiple testing using the Benjamini–Hochberg (BH) correction with a false discovery rate (FDR) ≤ 0.05 . Gene distribution was visualized with a volcano plot using the EnhancedVolcano function (EnhancedVolcano package)¹¹⁸. Genes with a Log₂Fold Change greater than 10 or lower than -10 were respectively considered to have a LFC equal to 10 or -10. Genes with a p -adjusted equal to “NA” were considered to have a p -adjusted equal to 1. Genes with a LFC equal to “NA” were considered to have an LFC equal to 0. Genes with a p -adjusted inferior to 10^{-10} were considered to have a p -adjusted equal to 10^{-10} . DEGs heatmap was performed with the Heatmap function (ComplexHeatmap package)¹¹⁹, clustering was performed using Euclidean distance and “average” as clustering method. To be able to identify features biological identities and the pathways they belong to, a functional annotation analysis was performed for all the comparisons and for feature list of interest with gprofiler2 package¹²⁰. Different databases were used to annotate the DEGs: Gene Ontology (Molecular Functions—MF, Biological Processes—BP, Cellular Component—CC), Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome, WikiPathways (WP), Transfac (TF), miRTarBase (MIRNA), Human Protein Atlas (HPA), CORUM (CORUM protein complexes), Human Phenotype Ontology (HP), RNA central. Functional annotation results were visualized with ggpubr package (ref. 121) via Balloon plots. The Gene Ontology network visualization was obtained with the cnetplot function (enrichplot package)¹²².

Analysis of patient data

To analyze patient overall survival, The Cancer Genome Atlas (TCGA) Pancancer Sarcoma dataset (accessed October 2024, $n = 253$)¹²³ was retrieved from cBioPortal (ref. 124), jointly with patient metadata. The analysis was performed in R/Bioconductor environment. Patients were categorized into high-expression and low-expression groups using the `surv_cutpoint` (survminer package)¹²⁵. The p -value was computed using the `Survdiff` function from the survival package, and the hazard ratio was determined using the `coxph` function from the same package. These results were visualized through Kaplan-Meier plots with the `ggsurvplot` function. Censoring was applied for cases starting at 60 months (equivalent to 5 years) onward, and patients who passed away after that point were treated as though they were still alive before that time. Survival plots were created using the `plot_surv_area` function (`contsurvplot`)¹²⁶. The signature score displayed in the survival plot was calculated by taking the trimmed mean at 0.2 for each gene across all samples. TCGA Pancancer data from all cancer datasets (accessed July 2024, $n = 10967$; ref. 123) was retrieved and analyzed as described above.

Statistics and reproducibility

Data was analyzed by Prism 8 (GraphPad). Results in bar blots are shown as mean \pm standard deviation. Unless differentially specified, differences between experimental groups were tested by student's t test. Results were considered significant at $p < 0.05$, and not significant (n.s.) for $p > 0.05$. Precise p -values are reported in figures; p -values below 0.0001 are indicated as $p < 0.0001$. Number of N (biological replicates), n (analyzed nuclei), as well as analyzed features can be found in figures and figure legends. For native metaphase analyses, “ n ” refers to number of analyzed metaphase spreads and “ $n\#$ ” to the total number of chromosomes. The original data values used to generate the quantification graphs are available in the Source Data file associated with this manuscript. No statistical method was used to pre-determine sample size; no data were excluded from the analyses; experiments were not randomized; investigators were not blinded to allocation during experiments and outcome assessment. Western blots, immunofluorescence images and EMSA data shown in figures are representative images obtained from individual experiments. Immunoprecipitation and protein pull-down experiments were carried out in duplicate. Tests used for bioinformatics analysis are described in the sections “Bioinformatics analysis – ChIP-Seq” and Bioinformatics analysis–RNA-seq”. For patient survival, Kaplan-Meier survival curves were generated and analyzed by Log-Rank test.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The sequencing data for ChIP-Seq and RNA-Seq have been deposited in NCBI's Gene Expression Omnibus with the GEO Series accession number GSE281893 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE281893>] and GSE281892 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE281892>], respectively. This study analyzed existing, publicly available data from the TCGA Research Network. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD060026 [<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PX060026>]. Requests for further information and resources should be directed to the corresponding authors. Source data are provided with this paper.

References

1. Crossley, M. P., Bocek, M. & Cimprich, K. A. R-loops as cellular regulators and genomic threats. *Mol. Cell* **73**, 398–411 (2019).

2. Castillo-Guzman, D. & Chédin, F. Defining R-loop classes and their contributions to genome instability. *DNA Repair* **106**, 103182 (2021).
3. Sanz, L. A. et al. Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. *Mol. Cell* **63**, 167–178 (2016).
4. García-Muse, T. & Aguilera, A. R loops: from physiological to pathological roles. *Cell* **179**, 604–618 (2019).
5. Petermann, E., Lan, L. & Zou, L. Sources, resolution and physiological relevance of R-loops and RNA–DNA hybrids. *Nat. Rev. Mol. Cell Biol.* **23**, 521–540 (2022).
6. Brickner, J. R., Garzon, J. L. & Cimprich, K. A. Walking a tightrope: The complex balancing act of R-loops in genome stability. *Mol. Cell* **82**, 2267–2297 (2022).
7. Li, X. & Manley, J. L. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**, 365–378 (2005).
8. Luna, R., Gómez-González, B. & Aguilera, A. RNA biogenesis and RNA metabolism factors as R-loop suppressors: a hidden role in genome integrity. *Genes Dev.* **38**, 504–527 (2024).
9. Wahba, L., Amon, J. D., Koshland, D. & Vuica-Ross, M. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol. Cell* **44**, 978–988 (2011).
10. Bayona-Feliu, A., Barroso, S., Muñoz, S. & Aguilera, A. The SWI/SNF chromatin remodeling complex helps resolve R-loop-mediated transcription-replication conflicts. *Nat. Genet.* **53**, 1050–1063 (2021).
11. Herrera-Moyano, E., Mergui, X., García-Rubio, M. L., Barroso, S. & Aguilera, A. The yeast and human FACT chromatin-reorganizing complexes solve R-loop-mediated transcription-replication conflicts. *Genes Dev.* **28**, 735–748 (2014).
12. Prendergast, L. et al. Resolution of R-loops by INO80 promotes DNA replication and maintains cancer cell proliferation and viability. *Nat. Commun.* **11**, 4534 (2020).
13. Nguyen, D. T. et al. The chromatin remodelling factor `<sc>ATR</sc>` suppresses R-loops in transcribed telomeric repeats. *EMBO Rep.* **18**, 914–928 (2017).
14. Tsai, S. et al. ARID1A regulates R-loop associated DNA replication stress. *PLoS Genet.* **17**, e1009238 (2021).
15. Saha, S. et al. Resolution of R-loops by topoisomerase III- β (TOP3B) in coordination with the DEAD-box helicase DDX5. *Cell Rep.* **40**, 111067 (2022).
16. Sessa, G. et al. BRCA2 promotes DNA-RNA hybrid resolution by DDX5 helicase at DNA breaks to facilitate their repair. *EMBO J.* **40**, e106018 (2021).
17. Hatchi, E. et al. BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Mol. Cell* **57**, 636–647 (2015).
18. Schwab, R. A. et al. The Fanconi anemia pathway maintains genome stability by coordinating replication and transcription. *Mol. Cell* **60**, 351–361 (2015).
19. García-Rubio, M. L. et al. The Fanconi anemia pathway protects genome integrity from R-loops. *PLoS Genet* **11**, e1005674 (2015).
20. Matos, D. A. et al. ATR protects the genome against R loops through a MUS81-triggered feedback loop. *Mol. Cell* **77**, 514–527 (2020).
21. Manzo, S. G. et al. DNA Topoisomerase I differentially modulates R-loops across the human genome. *Genome Biol.* **19**, 1–18 (2018).
22. Promonet, A. et al. Topoisomerase 1 prevents replication stress at R-loop-enriched transcription termination sites. *Nat. Commun.* **11**, 1–12 (2020).
23. Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I. & Chédin, F. R-Loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* <https://doi.org/10.1016/j.molcel.2012.01.017> (2012).

24. Ginno, P. A., Lim, Y. W., Lott, P. L., Korf, I. & Chédin, F. GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Res.* **23**, 1590–1600 (2013).
25. Gambelli, A., Ferrando, A., Boncristiani, C. & Schoeftner, S. Regulation and function of R-loops at repetitive elements. *Biochimie* <https://doi.org/10.1016/j.biochi.2023.08.013> (2023)
26. Nadel, J. et al. RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. *Epigenet. Chromatin* **8**, 46 (2015).
27. El Hage, A., Webb, S., Kerr, A. & Tollervy, D. Genome-wide distribution of RNA:DNA Hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genet.* **10**, (2014).
28. Zeng, C., Onoguchi, M. & Hamada, M. Association analysis of repetitive elements and R-loop formation across species. *Mob. DNA* **12**, 3 (2021).
29. Arora, R. et al. RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat. Commun.* **5**, 1–11 (2014).
30. Velichko, A. K. et al. Hypoosmotic stress induces R loop formation in nucleoli and ATR/ATM-dependent silencing of nucleolar transcription. *Nucleic Acids Res.* **47**, 6811–6825 (2019).
31. Shih, H.-T. et al. DNMT3b protects centromere integrity by restricting R-loop-mediated DNA damage. *Cell Death Dis.* **13**, 546 (2022).
32. Unoki, M. et al. CDCA7 and HELLS suppress DNA:RNA hybrid-associated DNA damage at pericentromeric repeats. *Sci. Rep.* **10**, 17865 (2020).
33. Lim, Y. W., Sanz, L. A., Xu, X., Hartono, S. R. & Chédin, F. Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi-Goutières syndrome. *eLife* **4**, e08007 (2015).
34. Yan, P. et al. Genome-wide R-loop Landscapes during Cell Differentiation and Reprogramming. *Cell Rep.* **32**, 107870 (2020).
35. Sagie, S. et al. Telomeres in ICF syndrome cells are vulnerable to DNA damage due to elevated DNA:RNA hybrids. *Nat. Commun.* **8**, 14015 (2017).
36. Killen, M. W., Stults, D. M., Adachi, N., Hanakahi, L. & Pierce, A. J. Loss of Bloom syndrome protein destabilizes human gene cluster architecture. *Hum. Mol. Genet.* **18**, 3417–3428 (2009).
37. He, Q. et al. The Daxx/Atrx complex protects tandem repetitive elements during DNA hypomethylation by promoting H3K9 trimethylation. *Cell Stem Cell* **17**, 273–286 (2015).
38. Yost, K. E. et al. Rapid and reversible suppression of ALT by DAXX in osteosarcoma cells. *Sci. Rep.* **9**, 1–11 (2019).
39. Pinto, L. M. et al. DAXX promotes centromeric stability independently of ATRX by preventing the accumulation of R-loop-induced DNA double-stranded breaks. *Nucleic Acids Res.* **52**, 1136–1155 (2024).
40. Liehr, T. Repetitive elements in humans. *Int. J. Mol. Sci.* **22**, 2072 (2021).
41. de Koning, A. P. J., Gu, W., Castoe, T. A., Batzer, M. A. & Pollock, D. D. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* **7**, e1002384 (2011).
42. Fox, A. H. & Lamond, A. I. Paraspeckles. *Cold Spring Harb. Perspect. Biol.* **2**, a000687 (2010).
43. Morozumi, Y., Takizawa, Y., Takaku, M. & Kurumizaka, H. Human PSF binds to RAD51 and modulates its homologous-pairing and strand-exchange activities. *Nucleic Acids Res.* **37**, 4296–4307 (2009).
44. Rajesh, C., Baker, D. K., Pierce, A. J. & Pittman, D. L. The splicing-factor related protein SFPQ/PSF interacts with RAD51D and is necessary for homology-directed repair and sister chromatid cohesion. *Nucleic Acids Res.* **39**, 132–145 (2011).
45. Sobinoff, A. P. et al. NONO, SFPQ, and PSCP1 promote telomerase recruitment to the telomere. *Nat. Commun.* **16**, 5769 (2025).
46. Petti, E. et al. SFPQ and NONO suppress RNA:DNA-hybrid-related telomere instability. *Nat. Commun.* **10** (2019).
47. Dong, L. et al. PTB-associated splicing factor (PSF) functions as a repressor of STAT6-mediated Igε gene transcription by recruitment of HDAC1*. *J. Biol. Chem.* **286**, 3451–3459 (2011).
48. Imamura, K. et al. Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. *Mol. Cell* **54**, 1055 (2014).
49. McCluggage, F. & Fox, A. H. Paraspeckle nuclear condensates: global sensors of cell stress? *Bioessays* **43**, e2000245 (2021).
50. Lee, M. et al. The structure of human SFPQ reveals a coiled-coil mediated polymer essential for functional aggregation in gene regulation. *Nucleic Acids Res.* **43**, 3826–3840 (2015).
51. Kameoka, S., Duque, P. & Konarska, M. M. p54(nrb) associates with the 5' splice site within large transcription/splicing complexes. *EMBO J.* **23**, 1782–1791 (2004).
52. Hall, L. L. et al. Stable COT-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell* **156**, 907–919 (2014).
53. Lukong, K. E., Huot, M.-E. & Richard, S. BRK phosphorylates PSF promoting its cytoplasmic localization and cell cycle arrest. *Cell Signal* **21**, 1415–1422 (2009).
54. Xiao, M. et al. Smad4 sequestered in SFPQ condensates prevents TGF-β tumor-suppressive signaling. *Dev. Cell* **59**, 48–63.e8 (2024).
55. Chakraborty, P., Huang, J. T. J. & Hiom, K. DHX9 helicase promotes R-loop formation in cells with impaired RNA splicing. *Nat. Commun.* **9**, 4346 (2018).
56. Rhee, D. K. et al. SFPQ, a multifunctional nuclear protein, regulates the transcription of PDE3A. *Biosci. Rep.* **37**, BSR20170975 (2017).
57. Marcassa, E. et al. Calpain mobilizes Atg9/Bif-1 vesicles from Golgi stacks upon autophagy induction by thapsigargin. *Biol. Open* **6**, 551–562 (2017).
58. Mahmud, I. & Liao, D. DAXX in cancer: phenomena, processes, mechanisms and regulation. *Nucleic Acids Res.* **47**, 7734–7752 (2019).
59. Lewis, P. W., Elsaesser, S. J., Noh, K.-M., Stadler, S. C. & Allis, C. D. Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proc. Natl. Acad. Sci.* <https://doi.org/10.1073/pnas.1008850107> (2010).
60. Lovejoy, C. A. et al. Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of telomeres pathway. *PLoS Genet.* **8**, e1002772 (2012).
61. Ueda, H. et al. Tumor suppressor functions of DAXX through histone H3.3/H3K9me3 pathway in pancreatic NETs. *Endocr.-Relat. Cancer* **25**, 619–631 (2018).
62. Kovac, M. et al. Exome sequencing of osteosarcoma reveals mutation signatures reminiscent of BRCA deficiency. *Nat. Commun.* **6**, 8940 (2015).
63. Kabeche, L., Nguyen, H. D., Buisson, R. & Zou, L. A mitosis-specific and R loop-driven ATR pathway promotes faithful chromosome segregation. *Science* **359**, 108–114 (2018).
64. Dionne, K. R., Zhuang, Y., Leser, J. S., Tyler, K. L. & Clarke, P. Daxx upregulation within the cytoplasm of reovirus-infected cells is mediated by interferon and contributes to apoptosis. *J. Virol.* **87**, 3447–3460 (2013).
65. Maiorino, L., Daßler-Plenker, J., Sun, L. & Egeblad, M. Innate immunity and cancer pathophysiology. *Annu Rev. Pathol.* **17**, 425–457 (2022).

66. Zhang, J., Yu, S., Peng, Q., Wang, P. & Fang, L. Emerging mechanisms and implications of cGAS-STING signaling in cancer immunotherapy strategies. *Cancer Biol. Med.* **21**, 45–64 (2024).
67. Britton, S. et al. DNA damage triggers SAF-A and RNA biogenesis factors exclusion from chromatin coupled to R-loops removal. *Nucleic Acids Res.* **42**, 9047–9062 (2014).
68. Deschamps, T. & Kalamvoki, M. Impaired STING pathway in human osteosarcoma U2OS cells contributes to the growth of ICPO-null mutant herpes simplex virus. *J. Virol.* **91**, e00006–e00017 (2017).
69. Basit, A. et al. The cGAS/STING/TBK1/IRF3 innate immunity pathway maintains chromosomal stability through regulation of p21 levels. *Exp. Mol. Med.* **52**, 643–657 (2020).
70. Wayne, J., Brooks, T., Landras, A. & Massey, A. J. Targeting DNA damage response pathways to activate the STING innate immune signaling pathway in human cancer cells. *FEBS J.* **288**, 4507–4540 (2021).
71. Yuan, W. et al. TDRD3 promotes DHX9 chromatin recruitment and R-loop resolution. *Nucleic Acids Res.* **49**, 8573–8591 (2021).
72. Liang, Z. et al. Binding of FANCI-FANCD2 complex to RNA and R-loops stimulates robust FANCD2 monoubiquitination. *Cell Rep.* **26**, 564–572.e5 (2019).
73. Yan, Q. et al. Proximity labeling identifies a repertoire of site-specific R-loop modulators. *Nat. Commun.* **13**, 53 (2022).
74. Elsässer, S. J., Noh, K.-M., Diaz, N., Allis, C. D. & Banaszynski, L. A. Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature* **522**, 240–244 (2015).
75. Goldberg, A. D. et al. Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140**, 678–691 (2010).
76. Udugama, M. et al. Histone variant H3.3 provides the heterochromatic H3 lysine 9 tri-methylation mark at telomeres. *Nucleic Acids Res.* **43**, 10227–10237 (2015).
77. Drané, P., Ouararhni, K., Depaux, A., Shuaib, M. & Hamiche, A. The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. *Genes Dev.* **24**, 1253–1265 (2010).
78. Groh, S. et al. Morc3 silences endogenous retroviruses by enabling Daxx-mediated histone H3.3 incorporation. *Nat. Commun.* **12**, 5996 (2021).
79. Truch, J. et al. The chromatin remodeller ATRX facilitates diverse nuclear processes, in a stochastic manner, in both heterochromatin and euchromatin. *Nat. Commun.* **13**, 3485 (2022).
80. Law, M. J. et al. ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell* **143**, 367–378 (2010).
81. McDowell, T. L. et al. Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes. *Proc. Natl. Acad. Sci. USA* **96**, 13983–13988 (1999).
82. Sadic, D. et al. Atrx promotes heterochromatin formation at retrotransposons. *EMBO Rep.* **16**, 836–850 (2015).
83. Voon, H. P. J. et al. ATRX plays a key role in maintaining silencing at interstitial heterochromatic loci and imprinted genes. *Cell Rep.* **11**, 405–418 (2015).
84. Wong, L. H. et al. ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. *Genome Res.* **20**, 351–360 (2010).
85. Clynes, D. et al. ATRX dysfunction induces replication defects in primary mouse cells. *PLoS One* **9**, e92915 (2014).
86. Clynes, D. et al. Suppression of the alternative lengthening of telomere pathway by the chromatin remodelling factor ATRX. *Nat. Commun.* <https://doi.org/10.1038/ncomms8538> (2015)
87. Ritchie, K. et al. Loss of ATRX leads to chromosome cohesion and congression defects. *J. Cell Biol.* **180**, 315–324 (2008).
88. De La Fuente, R., Viveiros, M. M., Wigglesworth, K. & Eppig, J. J. ATRX, a member of the SNF2 family of helicase/ATPases, is required for chromosome alignment and meiotic spindle organization in metaphase II stage mouse oocytes. *Dev. Biol.* **272**, 1–14 (2004).
89. Iwase, S. et al. ATRX ADD domain links an atypical histone methylation recognition mechanism to human mental-retardation syndrome. *Nat. Struct. Mol. Biol.* **18**, 769–776 (2011).
90. Hoelper, D., Huang, H., Jain, A. Y., Patel, D. J. & Lewis, P. W. Structural and mechanistic insights into ATRX-dependent and -independent functions of the histone chaperone DAXX. *Nat. Commun.* **8**, 1193 (2017).
91. Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G. & Raucher, F. J. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.* **16**, 919–932 (2002).
92. Sachs, P. et al. SMARCD1 ATPase activity is required to silence endogenous retroviruses in embryonic stem cells. *Nat. Commun.* **10**, 1335 (2019).
93. Gaurav, N. et al. The HP1 box of KAP1 organizes HP1 α for silencing of endogenous retroviral elements in embryonic stem cells. *Nat. Commun.* **16**, 5066 (2025).
94. Rowe, H. M. et al. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237–240 (2010).
95. Okamoto, Y. et al. FANCD2 protects genome stability by recruiting RNA processing enzymes to resolve R-loops during mild replication stress. *FEBS J.* **286**, 139–150 (2019).
96. Bhowmick, R. et al. RAD51 protects human cells from transcription-replication conflicts. *Mol. Cell* **82**, 3366–3381.e9 (2022).
97. Olazabal-Herrero, A. et al. The FANCI/FANCD2 complex links DNA damage response to R-loop regulation through SRSF1-mediated mRNA export. *Cell Rep.* **43**, 113610 (2024).
98. Baumann, C., Viveiros, M. M. & De La Fuente, R. Loss of maternal ATRX results in centromere instability and aneuploidy in the mammalian oocyte and pre-implantation embryo. *PLOS Genet.* **6**, e1001137 (2010).
99. Lin, C.-J., Conti, M. & Ramalho-Santos, M. Histone variant H3.3 maintains a decondensed chromatin state essential for mouse preimplantation development. *Development* **140**, 3624–3634 (2013).
100. Crossley, M. P. et al. R-loop-derived cytoplasmic RNA-DNA hybrids activate an immune response. *Nature* **613**, 187–194 (2023).
101. Izquierdo-Bouldstridge, A. et al. Histone H1 depletion triggers an interferon response in cancer cells via activation of heterochromatic repeats. *Nucleic Acids Res.* **45**, 11622–11642 (2017).
102. Marritt, K. L. et al. Intratumoral STING activation causes durable immunogenic tumor eradication in the KP soft tissue sarcoma model. *Front. Immunol.* **13**, 1087991 (2022).
103. Savitsky, P. et al. High-throughput production of human proteins for crystallization: the SGC experience. *J. Struct. Biol.* **172**, 3–13 (2010).
104. Dinami, R. et al. miR-155 drives telomere fragility in human breast cancer by targeting TRF1. *Cancer Res.* **74**, 4145–4156 (2014).
105. Bhatia, V. et al. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* **511**, 362–365 (2014).
106. Della Chiara, G. et al. Epigenomic landscape of human colorectal cancer unveils an aberrant core of pan-cancer enhancers orchestrated by YAP/TAZ. *Nat. Commun.* **12**, 2340 (2021).
107. Cawthon, R. M. Telomere length measurement by a novel monochrome multiplex quantitative PCR method. *Nucleic Acids Res.* **37**, 21 (2009).
108. Chakraborty, P. & Grosse, F. Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair* **10**, 654–665 (2011).

109. Stults, D. M., Killen, M. W. & Pierce, A. J. The sister chromatid exchange (SCE) assay. *Methods Mol. Biol.* **1105**, 439–455 (2014).
110. Johnson, W. L. et al. RNA-dependent stabilization of SUV39H1 at constitutive heterochromatin. *Elife* **6**, e25299 (2017).
111. Suryatenggara, J., Yong, K. J., Tenen, D. E., Tenen, D. G. & Bassal, M. A. ChIP-AP: an integrated analysis pipeline for unbiased ChIP-seq analysis. *Brief. Bioinform.* **23**, bbab537 (2022).
112. Huber, W. et al. Orchestrating high-throughput genomic analysis with bioconductor. *Nat. Methods* **12**, 115–121 (2015).
113. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comput Biol.* **9**, e1003118 (2013).
114. Babraham Bioinformatics—FastQC a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
115. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
116. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
117. ggplot2: Elegant Graphics for Data Analysis | SpringerLink. <https://link.springer.com/book/10.1007/978-3-319-24277-4>.
118. EnhancedVolcano: publication-ready volcano plots with enhanced colouring and labeling. <https://bioconductor.org/packages/devel/bioc/vignettes/EnhancedVolcano/inst/doc/EnhancedVolcano.html>.
119. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
120. Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* **35**, W193–W200 (2007).
121. Kassambara, A. Ggpubr 'Ggplot2' Based Publication Ready Plots. - References - Scientific Research Publishing. <https://www.scirp.org/reference/referencespapers?referenceid=3546076> (2022).
122. enrichplot. *Bioconductor* <http://bioconductor.org/packages/enrichplot/>.
123. The Cancer Genome Atlas Program (TCGA) - NCI. <https://www.cancer.gov/ccg/research/genome-sequencing/tcga> (2022).
124. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
125. Kassambara, A., Kosinski, M. & Biecek, P. Drawing Survival Curves using 'ggplot2' [R package survminer version 0.4.8]. in (2020).
126. Denz, R. & Timmesfeld, N. Visualizing the (Causal) effect of a continuous variable on a time-to-event outcome. *Epidemiology* **34**, 652–660 (2023).
- Curie grant agreement n. 859853. S.P. and G.C. acknowledge the International Center for Genetic Engineering and Biotechnology (ICGEB) for financial support. G.C. is a PhD student of the Molecular Biomedicine program and acknowledges support provided by University of Trieste. E.P. was supported by an AIRC fellowship 18026. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We are grateful to P. Calsou and S. Britton for providing U-2 OS cells containing an inducible system for RNaseH1 expression and M. Pagani and G. Della Chiara for support on the preparation of material for ChIP-Seq. We thank G. Del Sal for providing pLV-eGFP. The results presented here are in part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>.

Author contributions

Conceptualization, S.S and R.B. Methodology, A.F., M.G., L.M.R.N, G.C, A.F., A.G., P.V.B., A.Z, E.P., C.B., S.P., R.B. and S.S. Software, G.C., A.P. and S.P. Formal analysis, A.F., M.G., A.F., A.G., P.V.B., A.Z., L.M.R.N., G.C. and E.P. Investigation, A.F., M.G., A.F., A.P., A.G., P.V.B., A.Z., L.M.R.N., G.C. and E.P. Resources, S.P., S.O., R.B. and S.S. Data curation, S.P., L.M.R.N and S.S. Writing - original draft, R.B and S.S Visualization, A.F., A.G., P.V.B., A.Z., M.G., L.M.R.N. and G.C. Supervision, S.P., S.O., R.B., S.S. Project administration, S.S. Funding acquisition, S.S.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-026-69479-w>.

Correspondence and requests for materials should be addressed to Roberta Benetti or Stefan Schoeftner.

Peer review information *Nature Communications* thanks Jordi Bernues and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026

Acknowledgements

The laboratory of S.S. is funded by the Fondazione AIRC per la Ricerca sul Cancro Investigator grant IG 2019 Id.23074; the European Union - Next Generation EU, Mission 4, Component 2, CUP B93D21010860004; by the European Union - Next Generation EU, Mission 4, Component 1, Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) 2022, Prot. 2022KWLYA, CUP J53D23012780006; by the European Union - Next-GenerationEU. Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) 2022 PNRR Prot. P2022RSP2C. R.B. is supported by the European Union - Next Generation EU, Mission 4 Component 1, Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) 2022, Prot. 2022KWLYA, CUP G53D2300558006. S.O. is supported by Fondazione AIRC per la Ricerca sul Cancro Investigator grant IG20778 and the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-