LETTER

DHX9 suppresses RNA processing defects originating from the *Alu* invasion of the human genome

Tuğçe Aktaş[1]*, İbrahim Avşar Ilık[1]*, Daniel Maticzka[2], Vivek Bhardwaj[1,3], Cecilia Pessoa Rodrigues[1,3], Gerhard Mittler[1], Thomas Manke[1], Rolf Backofen[2] & Asifa Akhtar[1]

Transposable elements are viewed as 'selfish genetic elements', yet they contribute to gene regulation and genome evolution in diverse ways[1]. More than half of the human genome consists of transposable elements[2]. *Alu* elements belong to the short interspersed nuclear element (SINE) family of repetitive elements, and with over 1 million insertions they make up more than 10% of the human genome[2]. Despite their abundance and the potential evolutionary advantages they confer, *Alu* elements can be mutagenic to the host as they can act as splice acceptors, inhibit translation of mRNAs and cause genomic instability[3]. *Alu* elements are the main targets of the RNA-editing enzyme ADAR[4] and the formation of *Alu* exons is suppressed by the nuclear ribonucleoprotein HNRNPC[5], but the broad effect of massive secondary structures formed by inverted-repeat *Alu* elements on RNA processing in the nucleus remains unknown. Here we show that DHX9, an abundant[6] nuclear RNA helicase[7], binds specifically to inverted-repeat *Alu* elements that are transcribed as parts of genes. Loss of DHX9 leads to an increase in the number of circular-RNA-producing genes and amount of circular RNAs, translational repression of reporters containing inverted-repeat *Alu* elements, and transcriptional rewiring (the creation of mostly nonsensical novel connections between exons) of susceptible loci. Biochemical purifications of DHX9 identify the interferon-inducible isoform of ADAR (p150), but not the constitutively expressed ADAR isoform (p110), as an RNA-independent interaction partner. Co-depletion of ADAR and DHX9 augments the double-stranded RNA accumulation defects, leading to increased circular RNA production, revealing a functional link between these two enzymes. Our work uncovers an evolutionarily conserved function of DHX9. We propose that it acts as a nuclear RNA resolvase that neutralizes the immediate threat posed by transposon insertions and allows these elements to evolve as tools for the post-transcriptional regulation of gene expression.

Mammalian cells react to excessive double-stranded RNA (dsRNA) as a sign of viral infection with an intricate system collectively called the dsRNA response[8]. ADAR[3], PKR[9], Staufen[10], as well as cytoplasmic RNA helicases MDA-5 and RIG-I[11], are known to be involved in suppression or activation of the dsRNA response. DHX9 has a unique domain organization that resembles ADAR and PKR (Extended Data Fig. 1a). Surprisingly, despite its essential nature in mouse and human cells[12,13] and abundance (>1 million copies per cell[6]), little is known about the role of DHX9 in cellular homeostasis. Thus, in order to shed light on the cellular function of DHX9, we first identified its *in vivo* targets by using a UV-crosslinking-based method (uvCLAP: UV crosslinking and affinity purification, Extended Data Fig. 1b, see Methods). Generation of a cell line that expresses DHX9 with a 3×Flag and HBH tag at its C terminus[14] allowed direct comparison with other RNA-binding proteins including QKI-5, QKI-6, KHDRBS1-3, HNRNPK and EIF4A1 (Fig. 1a and Extended Data Figs 1, 2). We observed that human/mouse DHX9 and
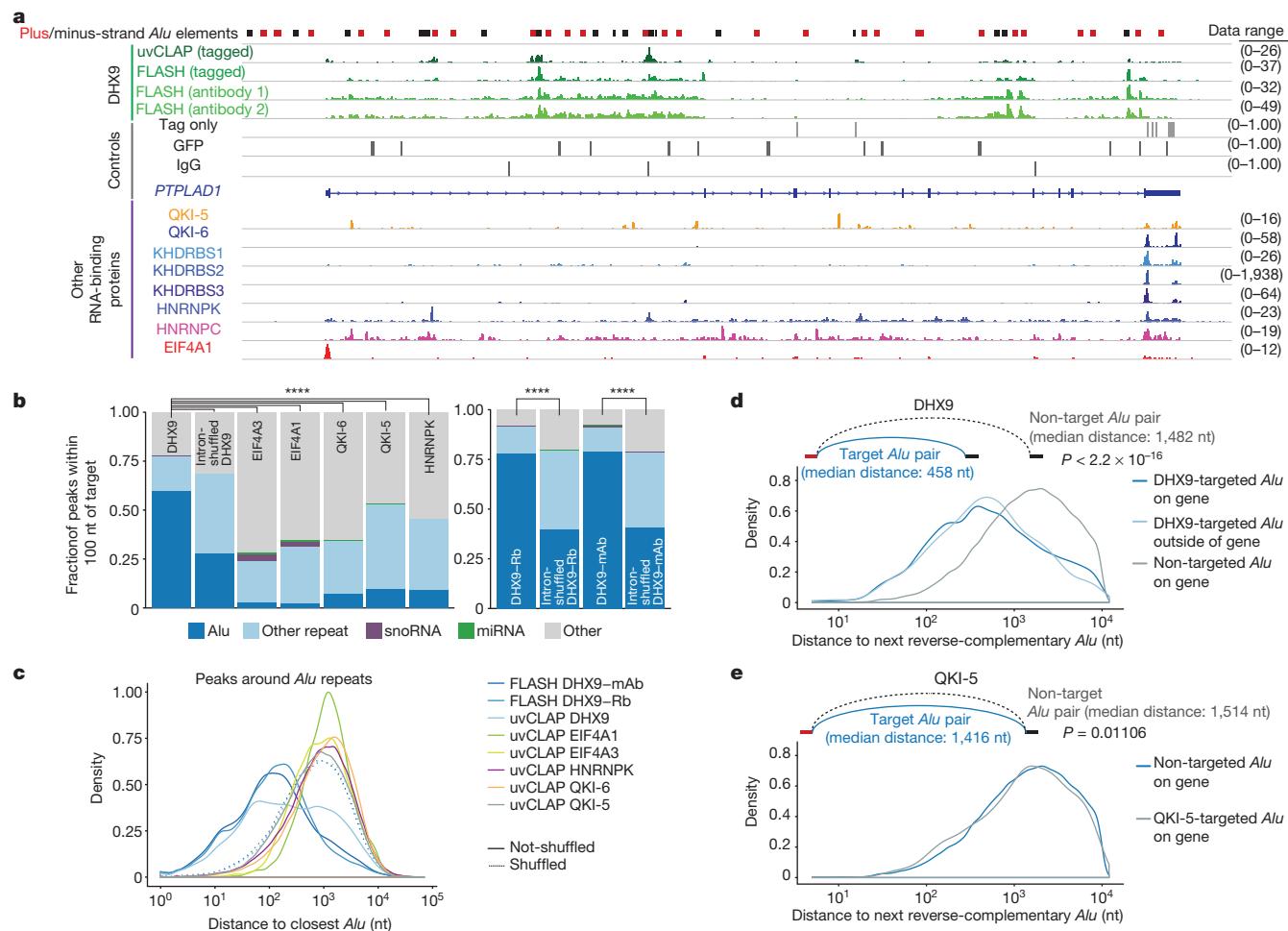
its fly orthologue Maleless (Mle) interacts mainly with intronic RNA (Extended Data Fig. 2i). Closer inspection revealed that DHX9 peaks are significantly enriched on *Alu* SINEs (Fig. 1a): 60% of DHX9 peaks are on or within 100 nucleotides (nt) of *Alu* repeats; compared to 3–10% for controls (Fig. 1b, all *P* values $<2.2 \times 10^{-16}$, Fisher's exact test). Using another UV-crosslinking method (FLASH: fast ligation of RNA after some sort of affinity purification for high-throughput sequencing, see Methods) that can utilize antibodies against endogenously expressed proteins in addition to tagged constructs, we obtained virtually identical DHX9 profiles to those derived with the tagged protein using uvCLAP, confirming the robustness of our observations (Fig. 1a, also see Extended Data Fig. 2a–d). DHX9 peaks obtained by FLASH fall predominantly on *Alu* repeats (Fig. 1b and Extended Data Fig. 1d) and shuffling DHX9 FLASH peaks within introns significantly reduces the percentage of peaks that fall on *Alu* elements (Fig. 1b). Analysis of uvCLAP/FLASH peaks that do not fall directly on *Alu* elements shows that DHX9 peaks tend to be closer to *Alu* elements compared to peaks of other RNA-binding proteins (Fig. 1c and Extended Data Fig. 1c). Shuffling the DHX9 peaks randomly within introns abrogates this closeness to *Alu* elements (Fig. 1c and Extended Data Fig. 1c). Using two genome-mapping-free approaches we further verified that the *Alu* enrichment we score is specific to DHX9 (see Methods and Extended Data Fig. 2e–g).

As *Alu* elements are strictly primate specific-retrotransposons[15], we next tagged endogenous DHX9 with 3×Flag–Avitag[16] in mouse embryonic stem cells and carried out uvCLAP/FLASH experiments (Extended Data Figs 2h and 3a–c). In mice, DHX9 peaks fall predominantly on B1 SINEs (Extended Data Fig. 1f and 3d–f). DHX9 peaks that do not directly fall on B1 elements are close to them compared to shuffled controls (Extended Data Fig. 1g, h). Notably, both *Alu* elements and B1 elements evolved independently from the ancestral 7SL RNA[17] (Extended Data Fig. 1e). Similar to the human data, we used genome-mapping-free methods to verify the specificity of B1 enrichment in DHX9 data (Extended Data Fig. 3d, e).

The common denominator between *Alu*, B1 and roX RNAs (main targets of human, mouse and *Drosophila* DHX9) is their propensity to form structured RNAs[18,19] that can directly attract DHX9 activity. Notably, however, even a visual inspection of DHX9 binding reveals that not all *Alu* elements are DHX9 targets (Fig. 1a). In order to determine the rules of DHX9 binding, we used a genomic similarity search tool (YASS[20]) to look for *Alu* elements that can base pair with each other to form exceptionally long dsRNA (Extended Data Fig. 4a, b). We observed that *Alu* elements that are not targeted by DHX9 have a median distance of 1,482 nt ($n = 173,653$) to their closest potential binding partner, whereas *Alu* elements that are targeted by DHX9 have a much shorter median distance of 458 nt ($n = 13,663$) indicating that DHX9 specifically targets long dsRNA formed by base pairing *Alu* elements (Fig. 1d, two-tailed Mann–Whitney *U*-test, $P < 2.2 \times 10^{-16}$;

Figure 1 | uvCLAP and FLASH reveal DHX9 binding on *Alu* elements.
a, IGV (Integrative Genomics Viewer) snapshot showing tagged (uvCLAP) and endogenous (FLASH) DHX9 binding on *Alu* elements of PTPLAD1. Additional tracks include control experiments and other RNA-binding proteins (tracks show merged biological duplicates). 'Data range' shows the coverage of uniquely mapped alignments for each profile.
b, Enrichment of uvCLAP (left) and FLASH (right) peaks on *Alu* (all *P* values $<2.2 \times 10^{-16}$, Fisher's exact test) or other elements in humans.

miRNA, microRNA; snoRNA, small nucleolar RNA. **c**, Distance of uvCLAP/FLASH peaks to the nearest *Alu* element. DHX9 peaks are closer to *Alu* elements, compared to shuffled peaks and other RNA-binding proteins (all *P* values $<2.2 \times 10^{-16}$, one-tailed Mann–Whitney *U*-test). **d**, **e**, Distance distributions of target, or non-target *Alu* pairs in DHX9 (**d**) or QKI-5 (**e**) uvCLAP peaks (*P* values calculated with two-tailed Mann–Whitney *U*-test, see Supplementary Table 3).

Extended Data Fig. 4c–h). DHX9 FLASH yields virtually identical results while other RNA-binding proteins appear to be agnostic to the positioning of *Alu* elements in and around their binding sites (Fig. 1e, Extended Data Fig. 4c–h and Supplementary Table 3).

Presence of *Alu* repeats and complementary sequences around exons correlate with circular RNA (circRNA) formation[21,22]. To determine whether DHX9 plays a role in this process, we first checked the abundance of previously reported circRNAs[23] using quantitative reverse-transcription PCR (RT–qPCR) in DHX9-depleted cells and observed a robust increase in circRNA levels (Extended Data Fig. 5a). Next, we generated sequencing libraries from poly(A)$^+$-depleted total RNA and indeed observed a clear, reproducible and global increase in the number of unique circRNAs in DHX9-depleted cells (Fig. 2a). In total, we detected 25,658 unique circRNAs in control short interfering RNA (siRNA)-treated samples and 50,355 in DHX9-siRNA-treated samples. DHX9-specific circRNAs contain significantly more *Alu* elements in flanking introns compared to control-specific circRNAs (90.65% in at least one flanking intron, 71.78% in both, $P < 2.2 \times 10^{-16}$, Fisher's exact test). Moreover, DHX9-depleted cells showed more than a fourfold increase in circRNA-producing genes compared to controls (Fig. 2b). We independently verified the upregulation and RNaseR insensitivity of the ten most upregulated circRNAs by RT–qPCR (Fig. 2c and Extended Data Fig. 5b).

In order to determine whether DHX9 has any effect on mRNAs that contain inverted-repeat *Alu* elements in their 3′ UTRs, we used a luciferase reporter system comparing constructs that have inverted-repeat-*Alu*-containing 3′ UTRs to *Alu*-free 3′ UTR constructs (Extended Data Fig. 5c, d). We observed a significant downregulation of luciferase activity for constructs that contain 3′ UTR inverted-repeat *Alu* elements upon *DHX9* depletion (Fig. 2d, Extended Data Fig. 5e–g). This effect could be rescued by overexpressing a wild-type DHX9 transgene but not a 'helicase-dead' mutant[24] (DHX9(GET)), suggesting that DHX9 resolves rather than passively coats inverted-repeat *Alu* elements, as previously suggested for Staufen proteins[25], in order to de-repress translation of mRNAs with inverted-repeat *Alu* elements in their 3′ UTRs (Fig. 2d, Extended Data Fig. 5c–g).
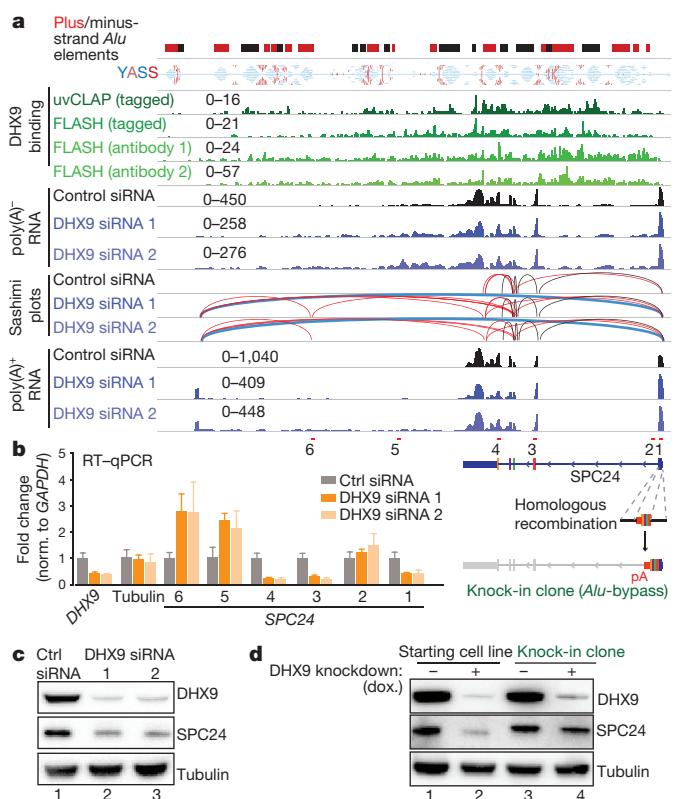
In addition to increasing circRNA formation, *Alu* elements can potentially disrupt normal RNA biogenesis by masking primary RNA sequences for splicing and transcript termination. Analysis of sequencing libraries prepared with poly(A)$^+$ RNA from DHX9-depleted cells revealed significant changes in the expression of 5,890 genes (Extended Data Fig. 6a–g). Further analysis showed 9,146 differentially spliced exons (at FDR-adjusted $P < 0.01$) affecting more than 4,400 genes in total (Extended Data Fig. 6e, f). Notably, CCL25, a cytokine that is almost exclusively expressed in the intestines and tissues involved in T-cell development[26] was among the most highly upregulated genes

Figure 2 | Loss of DHX9 protein causes increased circular RNA generation and impaired translation. a, Number of unique circRNAs detected in control (grey) or DHX9-siRNA-treated samples (orange). Error bars represent s.d. of biological quadruplicates. **b**, Number of genes producing circRNA species. Orange, DHX9-knockdown-specific; black, control-specific; grey, common. **c**, Top, RT–qPCR assays quantifying changes in linear RNAs versus circRNAs. Bottom, sensitivity of linear/ circRNA to RNaseR represented as the change in the value of crossing points (ΔCp). Error bars represent s.e.m. of four biological replicates. **d**, Luciferase assays show that DHX9 depletion significantly reduces expression of inverted-repeat-*Alu*-containing 3′ UTR elements, which can be alleviated with wild-type (WT) DHX9 overexpression and only partially with DHX9(GET). Mann–Whitney *U*-test was used to calculate *P* values. ****$P < 0.0001$; **$P = 0.0032$; wild-type DHX9, $P = 0.4278$. Error bars represent s.d. of two biological replicates.

Figure 3 | Loss of DHX9 leads to RNA processing defects. a, IGV plots of DHX9-binding (green) and poly(A)$^+$/poly(A)$^-$ RNA sequencing (blue) data and Sashimi plots depicting exon–exon connectivities (red, non-canonical splicing events; blue, connection between the first and the newly emerging terminal exon of *SPC24*). For visual clarity, only the reads on the same strand as *SPC24* are depicted. Data range shows the coverage of uniquely mapped alignments for each profile. **b**, Left, RT–qPCR validation of *SPC24* reduction following DHX9 depletion (qPCR amplicons (1–6) are depicted in **a**, bottom). Error bars represent s.e.m. of biological quadruplicates. Right, schematic representation of the Alu-bypass allele. **c**, SPC24 levels in HeLa cells upon DHX9 knockdown with two independent siRNAs compared to control siRNA. **d**, SPC24 levels are shown before (lanes 1, 3) or after (lanes 2, 4) DHX9 depletion, in parental (left) or *Alu*-bypass (right) cell line. Representative of three experiments.

(Extended Data Fig. 6h). Closer inspection of this locus revealed that upregulation of *CCL25* is probably caused by a splicing/termination failure during the processing of an *ELAVL1* anti-sense transcript (*ELAVL1^{AS}*) that is bound by DHX9 *in vivo* (Extended Data Fig. 6i). Sashimi plots show that upon DHX9 depletion, the first exon of *ELAVL1^{AS}* is spliced to a cryptic splice-acceptor site in the first exon of *CCL25* (approximately 40 kb downstream of *ELAVL1^{AS}*), leading to around a 60-fold upregulation of *CCL25*. Similar defects were observed in other highly upregulated genes (see Extended Data Fig. 7a–d).

We next investigated the consequences of DHX9 depletion on cell morphology and observed the formation of giant cells containing numerous small, pleomorphic nuclei (Extended Data Fig. 8a–c, Supplementary Videos 1–7). Live cell imaging revealed that cells lose the ability to align their chromosomes at the metaphase plate and eventually disintegrate into small and large spherical pieces that rapidly fuse to form a large cell containing a variable number of small nuclei (as many as 23 nuclei per cell, Extended Data Fig. 8c; also compare Supplementary Video 2 with Supplementary Video 7). A similar

phenotype has been reported by the MitoCheck consortium[27] under the 'Grape' category. Among the 153 Grape-phenotype-causing genes that were expressed in cells in our study, *SPC24* stands out as the most downregulated, and is one of four genes that has inverted *Alu* repeats at its 3′ UTR (Fig. 3a, b, left, and Extended Data Fig. 8d, left). Consistent with the chromosomal alignment defects, SPC24 is part of the Ndc80 complex, which acts as a tether between microtubules and chromosomes, forming an essential part of the kinetochore[28] (Extended Data Fig. 8d, right). Transcriptional reduction of *SPC24* also coincides with a reduction in its protein levels (Fig. 3c). Notably, both the 3′ UTR of *SPC24* and a downstream, *Alu*-rich intergenic domain are bound almost end-to-end by DHX9 (Fig. 3a). Similar to *CCL25*, Sashimi plots reveal that the first exon of *SPC24* is frequently spliced to a newly emerging exon around 15 kb downstream of the original terminal exon (blue lines in Fig. 3a). To rescue this defect, we created an '*Alu*-bypass allele' of *SPC24* by knocking-in a construct that contains the coding region of *SPC24* (amino acids 54–197), together with a polyadenylation site, immediately downstream of the first exon of *SCP24* that codes for the first 53 amino acids (Fig. 3b, right, and Extended Data Fig. 8e). Unlike the wild-type allele, the *Alu*-bypass allele is expressed independently of DHX9, indicating that the presence of *Alu* elements interferes with *SPC24* expression in the absence of DHX9 (Fig. 3d).

Figure 4 | RNA-independent interaction of ADAR^p150 with DHX9 is conserved between mouse and human. a, Silver-stained gel image for the tandem-affinity purification experiment (see Extended Data Fig. 9a–c). **b**, Immunoprecipitation of the endogenous DHX9 protein using formaldehyde-crosslinked nuclei or native whole-cell lysates. Lanes 1, 4: inputs for 0.5%, 0.1% formaldehyde-crosslinked nuclei, respectively; lanes 7, 10: inputs for the native lysates without and with RNaseA treatment, respectively; lanes 2, 5, 8, and 11: IgG controls; lanes 3, 6, 9, 12: DHX9 immunoprecipitations. For gel source data, see Supplementary Fig. 1. **c**, Mouse DHX9 interacts with nuclear ADAR(p150) (lanes 3 and 6) (Also see Extended Data Fig. 9h–l). **d**, DHX9 depletion leads to accumulation of dsRNA (green, DHX9; red, dsRNA (J2); blue, DNA (Hoechst). Representative of two independent experiments. **e**, RIP–qPCR (RNA immunoprecipitation quantitative PCR) with J2 antibody. Error bars show s.d. of two independent experiments.

Next, we used a stringent, SILAC (stable isotope labelling with amino acids in cell culture)-based purification strategy to determine the DHX9 interactome in human cells (Fig. 4a, Extended Data Fig. 9a–f and Methods). We identified 80 reproducible DHX9 interactors showing at least tenfold enrichment over background, (Supplementary Table 1) of which approximately 94% (75/80) are classified as 'RNA-binding' (UniProt[29]). Purifications performed under native conditions with or without RNase treatment revealed that the majority of these interactions are RNA-bridged (Extended Data Fig. 9d–f and Fig. 4b). Notably, while we detected two major isoforms of ADAR (ADAR(p150) and ADAR(p110); interferon-inducible and constitutive isoforms of ADAR, respectively) co-immunoprecipitating with DHX9 under cross-linked and native conditions, only ADAR(p150)-interaction resisted RNase-treatment (Fig. 4b, lanes 9 and 12). We further verified this interaction by expressing Flag-tagged ADAR(p150) in HEK293 cells (Extended Data Fig. 9g). Despite the evolutionary divergence of SINEs, the DHX9–ADAR(p150) interaction is conserved in mice (Fig. 4c) and is most pronounced in the nucleus, even though ADAR(p150) is equally distributed between the nucleus and cytoplasm (Extended Data Fig. 9h–l).

Finally, we observed that DHX9-depletion leads to appearance of J2 mAb-positive speckles, indicative of accumulating dsRNA in cells (Fig. 4d and Extended Data Fig. 10a). RNA immunoprecipitation–qPCR experiments using the same antibody revealed that DHX9 depletion leads to a two- to threefold increase in J2-recovered DHX9 targets, while ADAR depletion has little effect (Fig. 4e). Surprisingly, although we did not observe major changes in RNA-editing upon DHX9 depletion (Extended Data Fig. 10c), co-depletion of DHX9 and ADAR led to an augmented phenotype in all target regions both in immunoprecipitated dsRNA (Fig. 4e)

and circRNA production (Extended Data Figs 10b, d), suggesting a synergistic interaction.

In summary, DHX9 binds to independently evolved SINEs in humans and mice and interacts specifically with ADAR(p150), underscoring its role as a nuclear dsRNA resolvase both under normal conditions and probably during viral invasion. We propose that elevated levels of DHX9 as a potential pre-emptive measure against viral invasions can increase the tolerance of the host genome for a higher number of SINE insertions over the course of evolution which may enhance SINE/LINE-mediated genomic or transcriptomic innovation in evolutionarily complex organisms (Extended Data Fig. 10e).

1. Kazazian, H. H., Jr. Mobile elements: drivers of genome evolution. *Science* **303,** 1626–1632 (2004).
2. 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).
3. Elbarbary, R. A., Lucas, B. A. & Maquat, L. E. Retrotransposons as regulators of gene expression. *Science* **351,** aac7247 (2016).
4. Kim, D. D. Y. *et al.* Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Res.* **14,** 1719–1725 (2004).
5. Zarnack, K. *et al.* Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* **152,** 453–466 (2013).
6. Hein, M. Y. *et al.* A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* **163,** 712–723 (2015).

7. Koh, H. R., Xing, L., Kleiman, L. & Myong, S. Repetitive RNA unwinding by RNA helicase A facilitates RNA annealing. *Nucleic Acids Res*. **42,** 8556–8564 (2014).
8. Gantier, M. P. & Williams, B. R. G. The response of mammalian cells to double-stranded RNA. *Cytokine Growth Factor Rev*. **18,** 363–371 (2007).
9. Lemaire, P. A., Anderson, E., Lary, J. & Cole, J. L. Mechanism of PKR Activation by dsRNA. *J. Mol. Biol*. **381,** 351–360 (2008).
10. Elbarbary, R. A., Li, W., Tian, B. & Maquat, L. E. STAU1 binding 3′ UTR IR*Alu*s complements nuclear retention to protect cells from PKR-mediated translational shutdown. *Genes Dev*. **27,** 1495–1510 (2013).
11. Oshiumi, H., Kouwaki, T. & Seya, T. Accessory factors of cytoplasmic viral RNA sensors required for antiviral innate immune response. *Front. Immunol*. **7,** 200 (2016).
12. Lee, C. G. *et al*. RNA helicase A is essential for normal gastrulation. *Proc. Natl Acad. Sci. USA* **95,** 13709–13713 (1998).
13. Wang, T. *et al*. Identification and characterization of essential genes in the human genome. *Science* **350,** 1096–1101 (2015).
14. Tagwerker, C. *et al*. A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with *in vivo*cross-linking. *Mol. Cell. Proteomics* **5,** 737–748 (2006).
15. Batzer, M. A. & Deininger, P. L. Alu repeats and human genomic diversity. *Nat. Rev. Genet*. **3,** 370–379 (2002).
16. Kim, J., Cantor, A. B., Orkin, S. H. & Wang, J. Use of *in vivo* biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. *Nat. Protocols* **4,** 506–517 (2009).
17. Kramerov, D. A. & Vassetzky, N. S. Origin and evolution of SINEs in eukaryotic genomes. *Heredity* **107,** 487–495 (2011).
18. Ilik, I. A. *et al*. Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in *Drosophila*. *Mol. Cell* **51,** 156–173 (2013).
19. Quinn, J. J. *et al*. Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. *Nat. Biotechnol*. **32,** 933–940 (2014).
20. Noé, L. & Kucherov, G. YASS: enhancing the sensitivity of DNA similarity search. *Nucleic Acids Res*. **33,** W540–W543 (2005).
21. Jeck, W. R. *et al*. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19,** 141–157 (2013).
22. Zhang, X.-O. *et al*. Complementary sequence-mediated exon circularization. *Cell* **159,** 134–147 (2014).
23. Memczak, S. *et al*. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495,** 333–338 (2013).
24. Lee, C. G., Chang, K. A., Kuroda, M. I. & Hurwitz, J. The NTPase/helicase activities of *Drosophila* maleless, an essential factor in dosage compensation. *EMBO J*. **16,** 2671–2681 (1997).
25. Ricci, E. P. *et al*. Staufen1 senses overall transcript secondary structure to regulate translation. *Nat. Struct. Mol. Biol*. **21,** 26–35 (2014).
26. Svensson, M. & Agace, W. W. Role of CCL25/CCR9 in immune homeostasis and disease. *Expert Rev. Clin. Immunol*. **2,** 759–773 (2006).
27. Neumann, B. *et al*. Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* **464,** 721–727 (2010).
28. McCleland, M. L. *et al*. The highly conserved Ndc80 complex is required for kinetochore assembly, chromosome congression, and spindle checkpoint activity. *Genes Dev*. **17,** 101–114 (2003).
29. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res*. **43,** D204–D212 (2015).

Supplementary Information is available in the online version of the paper.

Author Contributions T.A., I.A.I. and A.A. designed the experiments, T.A. and I.A.I. performed most of the experiments except live-cell imaging and mass-spectrometry, which were carried out by C.P.R. and G.M., respectively; I.A.I. and T.A. developed uvCLAP and FLASH methods with input from D.M., R.B. and A.A.; D.M. analysed uvCLAP and FLASH data and was supervised by R.B.; V.B. analysed the RNA-seq data, carried out repeat-enrichment and circRNA analysis and was supervised by T.M. and A.A.; I.A.I., T.A. and A.A. wrote the manuscript with input from D.M. and V.B.; R.B. and A.A. acquired funding; A.A. supervised all aspects of the study. All authors reviewed, edited and approved the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to A.A. (akhtar@ie-freiburg.mpg.de).

METHODS

No statistical methods were used to predetermine sample size. $P < 0.01$ was considered statistically significant. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Cell culture and generation of stable cell lines. HEK293FT (gift from the Jenuwein laboratory), HeLa (gift from the Ellenberg laboratory) and FLPin Trex HEK293 (Thermo Fisher Scientific, R78007) cells are maintained with DMEM-Glutamax supplemented with sodium pyruvate, glucose and 10% FBS. In addition to those supplements, FLPin Trex HEK293 cells were maintained in zeocin- and blasticidin-containing medium according to manufacturer's protocol and the zeocin selection is exchanged with hygromycin upon transgene transfection. All the transgenes were cloned into pCDNA5-FRT/To (Thermo Fisher Scientific catalogue no. V6520-20) with a C-terminal $3\times$Flag–HBH tag and were co-transfected with pOG44 plasmid with a 1:9 DNA concentration ratio. Cells were re-plated in different dilutions 24 h after transfection and $150\,\mu g\,ml^{-1}$ hygromycin selection was initiated 48 h after transfection. Cell lines were maintained with blasticidin and hygromycin at all times and the transgenes were induced with either with $0.1\,\mu g\,ml^{-1}$ (for uvCLAP) or $1.0\,\mu g\,ml^{-1}$ (for purifications) doxycyclin.

CRISPR–Cas9 facilitated endogenous GFP tagging of the human DHX9 was performed in FLPinTrex HEK293 cell line carrying the inducible SPOP-GFP-nanobody[30]. The repair template was co-transfected in a 3:1 ratio with SpCas9 (pX459, Addgene no. 62988) vector carrying the guide RNA #1 (Extended Data Fig. 5d). Cells were selected 48 h after the initial selection with $1\,\mu g\,ml^{-1}$ puromycin for 4 days total. Single-cell sorting into 96-well plates was carried out by gating for the cells with the strongest GFP signals (9.42% of all events). Colonies were expanded and screened for homozygously tagged DHX9 alleles using western blots. The knock-in of resistant allele at the SPC24 locus was performed in a homozygously GFP-tagged DHX9 clone using same conditions as described above (see Supplementary Table 2 for the sequences of guide RNAs, repair templates and screening oligos).

Mouse ES cells were maintained with 15% FBS, $2,000\,U\,ml^{-1}$ LIF, sodium pyruvate, NEAA, Glutamax and $0.1\,nM$ β-mercaptoethanol supplemented into DMEM. CRISPR–Cas9 facilitated endogenous tagging of the mouse DHX9 was performed in a mouse ES cell line (WT26 male ES cell line was a gift from the Jenuwein laboratory) that we modified by adding a BirA ligase (plasmid provided by the Orkin laboratory[16]). Endogenous tagging was achieved by a co-transfection of ssODN (sequences of ssODN and screening oligos are listed in Supplementary Table 2) that provided the repair template with a $3\times$Flag–Avitag tag and the Cas9 plasmid that provided the guide RNA (based on pX459; Zhang laboratory) according to an earlier published protocol[31]. Initial screening of the clones was performed by PCR amplification flanking the tagged locus and scoring a size shift in PCR product. Positive clones from the first screen are then validated for the expression of Flag-tagged DHX9 protein (see Supplementary Table 2 for the sequences of guide RNAs, repair templates and screening oligos). ES cells were differentiated into NPCs as described previously[32]. FLP-in-3T3 cells were purchased from Thermo Fisher (R76107) and maintained in DMEM-Glutamax supplemented with sodium pyruvate, glucose and 10% FBS and zeocin according to the manufacturer's protocol. All cell lines are regularly checked for the absence of mycoplasma by PCR detection kit (Jena Bioscience PP-401).

uvCLAP. Stable cell line derived from FLPin Trex HEK293 cells are induced with $0.1\,\mu g\,ml^{-1}$ doxycycline for $\sim$16 h; as controls, cell lines expressing either the tag alone or eGFP with the same tag were used. Cells are rinsed with PBS and crosslinked with $0.15\,mJ\,cm^{-2}$ UVC light. After crosslinking, cells are pelleted by centrifugation, snap-frozen in liquid nitrogen and kept at $-80\,°C$ until use. Cells are lysed with 0.5 ml of lysis buffer (50 mM Tris-Cl, pH 7.4; 140 mM NaCl, 1 mM EDTA, 1% Igepal CA-630, 0.1% SDS, 0.1% DOC), mildly sonicated and immunoprecipitated with anti-Flag beads for 1 h at 4 °C. Beads are washed with lysis buffer and bound material is eluted with $3\times$Flag peptide ($250\,\mu g\,ml^{-1}$). The eluate is then incubated with MyONEC1 beads to collect biotinylated target protein, after which the beads are washed with high-stringency buffers (0.1% SDS, 1 M NaCl, 0.5% LiDS, 0.5 M LiCl and 1% SDS, 0.5 M LiCl) to remove non-specific interactors. 3′ linkers are ligated with T4 RNA Ligase 1, excess adapters are washed away and 3′-tagged, cross-linked RNA is released with proteinase K digestion and column purification (Zymo DNA Clean & Concentrator). Reverse transcription is carried out with SuperScript III and barcoded reverse-transcription primers. After reverse transcription, relevant samples are mixed and the cDNA is separated on a 6% 6M Urea PAA gel. Size-fractionated cDNA is then processed essentially as described in the iCLIP protocol to generate sequencing libraries[33].

uvCLAP employs a tri-barcode approach that combines barcoded ScriptSeq Index PCR Primers and two custom barcodes located adjacent to the 5′ and 3′ adapters. Barcodes located at the 5′ end of the inserts were designed using edittag[34].

Barcodes located at the 3′ end of the inserts were created according to the semi-random patterns DRYYR and DYRRY (where D = not C; R = purine; Y = pyrimidine) and allowed to encode two mutually exclusive experimental conditions. uvCLAP uses a random tag strategy[35] to determine individual crosslinking events. uvCLAP primers contain 5 random nucleotides that are interleaved with the 5′ barcodes according to the pattern $NNNT_1T_2T_3T_4T_5NN$ (N = random nucleotide; T = barcode nucleotide). In combination with the semi-random 3′ barcodes, this allows us to distinguish at least $49,152\,(4^5 \cdot 3 \cdot 2^4)$ crosslinking events per nucleotide.

Size fractions were tagged via indexed PCR primers, pull-down conditions and biological replicates were tagged via 5′ and 3′ barcodes. Barcodes and random tags were extracted using custom scripts. Adapters were trimmed and the library was demultiplexed using Flexbar (v2.32)[36]. Possible readthroughs into the barcoded regions were removed by clipping 10 and 5 nt from the 3′ ends of first and second mate reads. Reads were mapped to the reference genomes (hg19, mm10, dm3) using bowtie2 (v2.2.0)[37] with parameters: –very-sensitive –end-to-end –no-mixed –no-discordant –maxins 200. We excluded all reads for which bowtie2 could identify multiple distinct alignments as indicated by the 'XS:i' flag and used the alignments of the remaining uniquely mapped reads to determine crosslinking events and crosslinked nucleotides as previously described[18]. Crosslinking events of the size fractions and biological replicates were combined.

uvCLAP peaks were called on crosslinked nucleotides with JAMM (version 1.0.7rev1, parameters: -d y -t paired -b 50 -w 1 -m normal)[38] using the two replicates of the respective pull-down condition as foreground and the two replicates of the corresponding control pull-down condition as background (Extended Data Fig. 2a–d, h). The peaks of all experiments were merged and used to calculate pairwise Spearman correlations among all replicates based on the number of events falling on each peak region, showing increased pairwise correlations between biological replicates within each condition over replicates of differing experimental conditions (deeptools version 2.3.5)[39] (see Extended Data Fig. 2a–d).

FLASH. Cells were crosslinked with $0.15\,mJ\,cm^{-2}$ UVC irradiation, lysed with $1\times$ NLB buffer ($1\times$ PBS, 0.3 M NaCl, 1% Triton-X, 0.1% Tween-20) and homogenized by water-bath sonication. The target protein of interest was pulled-down with paramagnetic beads pre-coupled to antibodies against the protein of interest. After a brief RNase I-digestion, RNA 3′ ends were healed with T4 PNK. Custom-made, barcoded adapters were ligated using T4 RNA Ligase 1 or T4 RNA Ligase 2KQ (if pre-adenylated adapters are used) for 1 h at 25 °C. Custom FLASH adapters contained two barcodes and random nucleotides adjacent to the 3′-adapters according to the pattern $NNB_1B_2NT_1T_2T_3T_4T_5T_6NN$ (N = random tag nucleotide; T = tag nucleotide; B = RY-space tag nucleotide). Random tags were used to merge PCR-duplicates, regular tags were used to specify the pull-down condition, and semi-random RY-space tags were used to distinguish the biological replicates. Excess adapters were washed away, negative controls were mixed with experimental controls and RNA was isolated with Proteinase K treatment and column purification. Isolated RNA was reverse-transcribed and treated with RNase H. cDNA was column-purified and circularized with CircLigase for 2–16 h. Circularized cDNA was directly PCR amplified, quantified with Qubit/Bioanalyzer and sequenced on Illumina NextSeq 500 in paired-end mode.

FLASH read processing and mapping was performed using the Galaxy platform[40]. Adapters were trimmed using Flexbar (v2.5). Libraries were demultiplexed using bctools (https://github.com/dmaticzka/bctools, v0.2.0) and Flexbar (v2.5). Possible readthroughs into the barcoded regions were removed by clipping 13 nt from the 3′ ends of first mate reads. Reads were mapped to the reference genomes (hg19, mm10) using bowtie2 (v2.2.6) with parameters: –very-sensitive –end-to-end –no-mixed –no-discordant –maxins 500. We excluded all reads for which bowtie2 could identify multiple distinct alignments as indicated by the XS:i flag and used the alignments of the remaining uniquely mapped reads to determine crosslinking events as previously described[18].

FLASH peaks were called on alignments of crosslinking events with PEAKachu (https://github.com/tbischler/PEAKachu/releases/tag/0.0.1alpha2, version 0.0.1alpha2, parameters: –pairwise_replicates -m 0 -n manual –size_factors 1 1 0.75 0.75)[41], using the two replicates of the respective pulldown condition as foreground and the two replicates of the corresponding control pull-down condition as background (Supplementary Table 3). Pairwise Spearman correlations were calculated as for uvCLAP (Extended Data Fig. 2a–d, h).

Graph based clustering and repeat enrichment of uvCLAP data. We followed a strategy described before to assess the enrichment of different repeat families in uvCLAP samples[42]. Briefly, we created a library of canonical repeat sequences and repeat instances from the corresponding human (hg38) and mouse (mm10) genome, from repeatmasker (RepeatMasker open-4.0.5). We then mapped the sequencing reads to this library, and the uniquely mapping reads were used to calculate the maximum likelihood estimate of enrichment (MLE) for each repeat family, for the test (eg. DHX9) samples over the corresponding control (empty vector)

sample. We then performed hierarchical clustering using the MLE scores to visualize uniquely enriched repeat families for each sample.

As an alternative approach, we followed a mapping-free, graph-based sequence clustering approach described previously for plant genomes[43]. In brief, we first sampled 100,000 reads from DHX9, EIF4A3, QKI-6 and QKI-5 datasets, followed by a clustering pipeline[44] which performs an all-to-all pairwise comparison between the reads in order to construct a graph, containing reads as nodes and the overlap between reads as edges. The vertices in the graph correspond to the reads and the edge weights represent the similarity score between the reads. The graph is then divided into clusters representing different connected communities. We then annotated the reads belonging to each cluster using RepeatMasker for the presence of different repeat families. Top clusters (with highest number of reads) for each sample are then visualized to assess the composition of major *Alu* families and other repeats. Increasing the sample size or multiple rounds of sampling had no effect on the composition of top clusters. Analysis code is publicly available at: https://github.com/vivekbhr/dhx-alu.

Detection of paired *Alu* elements. We performed a similarity search (YASS[20] v1.4) to determine pairs of sequences reverse-complementary to each other within each gene (Ensembl 75). For each experiment, we selected all genes containing at least one peak and three annotated *Alu* elements (UCSC RepeatMasker track). Within the selected genes we determined *Alu* pairs based on all reverse-complementary YASS pairs with distances below 10,000 nt. Two *Alu* elements were considered paired if at least 75% of the first *Alu* was overlapped by one part of the YASS alignment and at least 75% of the second *Alu* was overlapped by the corresponding reverse-complementary part of the YASS alignment. For each paired *Alu* we then determined the distance to its nearest partner. *Alu* elements were categorized as targeted if a peak was located within 100 nt, otherwise as untargeted. For DHX9 we repeated this analysis for targeted *Alu* elements in extragenic regions.

Sample preparation for qPCR detection of circular RNAs and RNA-seq libraries. The siRNA knockdowns (5 nM of Silencer Select (Ambion) control siRNA 4390846 or s4019 and s4020 siRNA for DHX9 and s1009 for ADAR) were performed on FLPinTrex HEK293 cells for 3 days in quadruplicates using RNAiMAX (Thermo Fisher Scientific). Total RNA extracts from siRNA transfected cells was prepared by Zymo Quick-RNA Kit and first-strand cDNA synthesis was carried out with SuperScriptIII. Quantitative real-time PCR is performed using the oligos listed in Supplementary Table 2 with FastStart SYBR Green Master (Roche 04 913 914 001). For RNase R treated samples: 500 ng of total RNA was incubated with 20 U of RNase R (Epicentre/Illumina) for 30 min at 30 °C after which the reactions were cleaned up with RNA Clean and Concentrator columns (Zymo) before proceeding with reverse transcription. For the poly(A)+ libraries, TruSeq Stranded mRNA Library Prep Kit (Illumina) was used as per manufacturer's instructions. For the poly(A)− libraries, poly(A)+ RNA was depleted using paramagnetic oligo-d(T) beads (Thermo Fisher Scientific) twice. The remaining, poly(A)+-depleted RNA is processed with the TruSeq Stranded Total RNA Library Prep Kit (Illumina).

RNA sequencing and analysis. Both poly(A)+ and poly(A)− reads were trimmed using Trim Galore to remove adaptors and mapped to GRCh38 (hg38) genome sequence using TopHat2 (v2.0.13)[45]. Poly(A) reads were assigned to features (Gencode 21) using featurecounts (v1.5.1) (counting properly paired primary alignments above mapping quality of 10)[46], and differential expression analysis was performed using edgeR (v3.12.0)[47]. We identified genes as differentially expressed if they are called significant in two independent comparisons (control vs siRNA1-s4019 and control vs siRNA2-s4020, FDR < 0.05), with expression fold change in same direction, in both comparisons. For analysis of differentially spliced exons, the gene models were flattened and differential expression was performed at exon level using DEXSeq (v1.18)[48]. Exons were called differentially spliced at FDR adjusted $P < 0.01$. For the poly(A)− library, the reads were mapped to the genome using STAR (v2.4.2)[49] (using option –chimSegmentMin 20) and the circular RNAs were annotated using circExplorer (v2.1)[22]. Depth of sequencing was not significantly different between samples, $P = 0.9482$, Welch t-test.

Luciferase assays. The siRNA knock-downs (5 nM of Silencer Select (Ambion) control siRNA 4390846 or s4019 siRNA for DHX9) were performed on HEK293FT or HeLa cells and the reporter constructs (50 ng of pmirGlo with 3′ UTR inserts) were transfected on the 48th hour of knockdown. All luciferase assays were performed on the 72nd hour of knockdown. The knockout clone was generated by co-transfecting the two guide RNAs; 1 and 2 (Extended Data Fig. 5d) co-expressing wild-type SpCas9 (pX459, Addgene no. 62988) and selected with 1 μg ml−1 puromycin for 4 days. Colonies were picked, screened and expanded. Rescue constructs (GFP, wild-type DHX9, and DHX9(GET)) were transfected into wild-type (only Cas9-expressing) and knockout clones and the transfection of the luciferase constructs was performed 24 h after the rescue construct transfection in technical quadruplicates (see Extended Data Fig. 5e for the expression validation of the rescue constructs). All luciferase reads were performed on the 48th hour of the

primary transfection. The data points in Fig. 2d (and in Extended Data Fig. 5f) show all technical quadruplicates of the two biological replicates (that is, eight data points per 3′ UTR construct).

Live cell imaging. All live cell imaging experiments were performed using monoclonal reporter cell lines that were generated as previously described[50]. In brief, HeLa 'Kyoto' cell line was transfected with pH2V-mpH2B-mCherry-IRES-neo3 and pmEGFP-α-tubulin-IRES-puro2b plasmids. Thus, in the presence of neomycin and puromycin cells are H2B-mCherry-tubulin-EGFP stable reporter cells. Cells were cultured in Dulbecco's modified eagle medium (DMEM; GIBCO) supplemented with 10% (v/v) and treated either with siRNA control or siRNA against DHX9 in eight-well coverslip-bottomed dishes (ibidi GmbH cat. no. 80826). Six hours following transfection, medium was changed and supplemented with 1% penicillin/streptomycin, puromycin (1 μg μl−1) and geneticin/neomycin (200 μg ml−1) (Invitrogen, Life Technologies, cat. no. 21810031). 24, 48 or 72 h post transfection, dishes were transferred to a Tokai Hit stage incubation unit, wherein cells were maintained at 37 °C and humidified atmosphere of 5% CO_2 throughout the experiment. Cells were imaged using a Zeiss Observer.Z1 with the CSU-X1 spinning disk head (Yokogawa) and the QuantEM:512SC camera (Photometrics) and were visualized using an air 40× objective. Z-stack parameters were taken every 5 min for at least 16 h and a maximum of 24 h. Laser power and exposure times were kept to a minimum condition (<9%). For analysis cells were manually counted and evaluated at the latest time point using Zen lite-Blue software (ZEISS). Nuclei numbers (Extended Data Fig. 8c) come from 15 fields of each condition at 24 h knockdown (in total 21 cells for wild type, 21 cells for DHX9 knockdown are included in the plot), from 5 fields for wild type and 15 fields for knockdown at 48 h knock-down (in total 28 cells for wild type and 40 cells for DHX9 knockdown are included in the plot), and from 10 fields for wild type and 15 fields for DHX9 knockdown at 72 h knockdown (in total 30 cells for wild type, 32 cells for DHX9 knockdown are included in the plot).

SILAC labelling of the cells and the biochemical purification of DHX9. Cells were adapted to SILAC medium (with K8R10 (1:1500 each) for Heavy or with K0R0 for Light) for 7 days and then were expanded to larger cell culture plates for another 7–10 days. Crosslinked purification was performed on 5–10 mg of 0.5% formaldehyde crosslinked nuclear extracts that were solubilized by mechanical shearing using sonication. Flag IP (Flag-M2 Agarose resin) was performed overnight at 4 °C and the proteins were eluted by 6 M urea-containing buffer (6M Urea, 0.5 M NaCl, 0.5% NP-40, 50 mM Na_2HPO_4, 50 mM Tris pH 8.0, 0.05% Tween-20). Streptavidin (using MyOneC1 magnetic beads) pull-down was performed on the diluted (down to 1 M urea) eluate for 3 h at 4 °C. Streptavidin-bound proteins were sequentially washed with buffers containing 1 M and 6 M Urea. For the in-gel analysis proteins were eluted from beads by boiling and for the in-solution analysis beads were rinsed three times in N-octyl-β-glucopyranoside containing Tris (pH 7.4) buffer before trypsinization. Native purifications were performed on whole-cell extracts that were prepared in 450 mM NaCl and 1% Triton-X containing buffer. Silver Gel staining was performed using the Silver Quest Silver Staining Kit (Thermo Fisher LC6070).

Immunoprecipitation. Native whole-cell extracts or crosslinked nuclear extracts were incubated with 1 μg of DHX9 (polyclonal rabbit, see 'Antibodies' section) or of IgG (rabbit Santa Cruz sc-2017) antibody overnight at 4 °C and coupled to the ProtG Dynabeads (Life Technologies 10004D) for 1 h. Beads were washed in the IP buffer three times for 5 min each. Elution from the beads was performed in 2× protein loading dye by incubating for 10 min at 70 °C (or 95 °C for crosslinked samples) with shaking. DHX9 was detected with monoclonal mouse antibody (for the rest of the detected proteins see 'Antibodies' section).

Immunofluorescence. Cells were crosslinked with 4% methanol-free formaldehyde in PBS at room temperature for 10 min and permeabilized with 0.1% Triton-X and 1% BSA in PBS for 30 min at room temperature. Primary antibodies (see details in the Antibody section) were diluted (1:1000 for the DHX9 pRb, 1:500 for the SC-35 (SRSF2) and J2 antibodies) in PBS with 0.1% Triton-X and 1% BSA and incubated with fixed cells at 4 °C for ~16 h. Fluorescently labelled secondary antibodies with the appropriate serotype were used reveal target proteins. Hoechst 33342 was used to stain DNA. Imaging was performed with a Leica SP5 confocal microscope.
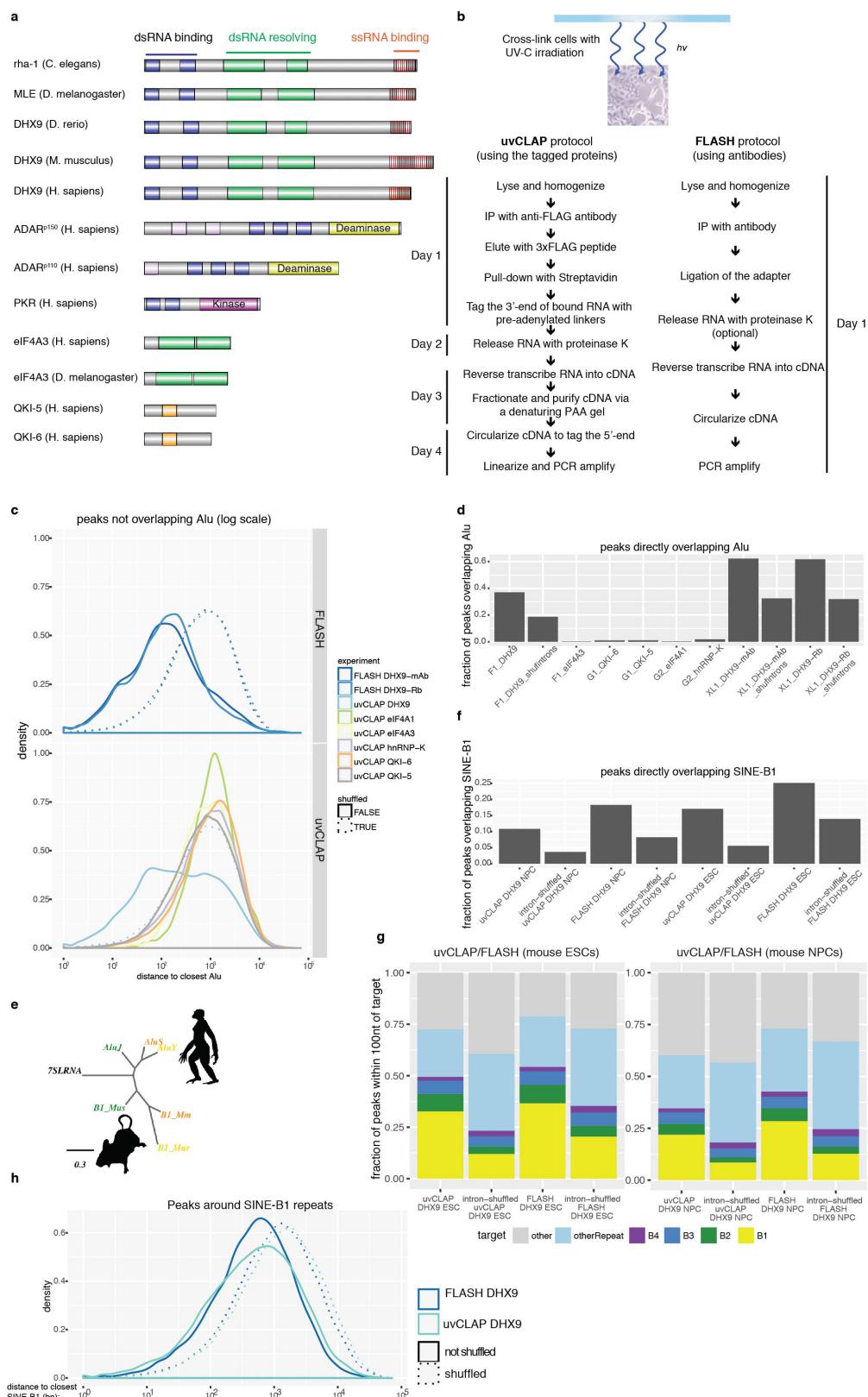
J2 (dsRNA) pull-down. Cells were harvested on the third day of siRNA treatment (siRNA transfections were carried out as described above) in RIP buffer (25 mM HEPES pH 7.2, 75 mM NaCl, 5 mM $MgCl_2$, 0.1% Igepal CA-630, 1 U μl−1 RNasin), sonicated (Bioruptor, 5 min, 30 s ON, 30 s OFF) and centrifuged at 20,000g for 20 min at 4 °C. Clarified lysates were incubated with 25 μl of ProtG Dynabeads pre-coupled to 2 μg of J2 mAb (Scicons) or mouse-IgG (Santa Cruz) for 3 h at 4 °C. The beads were washed three times with the RIP buffer and bound RNA was isolated using the RNA Clean and Concentrator columns (Zymo). Eluted RNA was reverse transcribed using SuperScript III with the following modifications:

RNA was incubated at 70 °C for 3 min to remove secondary structures, which is followed by reverse transcription at 50 °C for 15 min, 60 °C for 10 min and 70 °C for another 10 min. Quantitative real-time PCR was performed using the oligos listed in Supplementary Table 2 with FastStart SYBR Green Master (Roche 04 913 914 001), enrichments were calculated using the $2^{-\Delta\Delta C_t}$ method using 10% of input material as a reference for each immunoprecipitation.

Antibodies. DHX9 (monoclonal mouse, SantaCruz sc13183, polyclonal rabbit Abcam ab26271, monoclonal rabbit Abcam ab183731), ADAR (for the detection of human protein: Sigma SAB1401004, for the detection of mouse protein: Santa Cruz sc-73408), J2 antibody (Scicons, English and Scientific Consulting, Hungary), NONO (Sigma AV40715), ZFR (Sigma SAB2104153), HNRNPC (Santa Cruz sc-32308), KHDRBS1 (Sigma S9575), POL2 (4H8 Active Motif 101307, and 8WG16 Abcam ab817), CDK9 (Santa Cruz sc484), Flag-HRP (Sigma A8592), streptavidin-HRP (Pierce 21130), SPC24 (Bethyl, A304-260A), SRSF2/SC35 (Sigma S4045), MSL2 (Sigma HPA003413), MSL1 (polyclonal rabbit antibody generated by the Akhtar laboratory, previously described in ref. 32), MOF (Bethyl A300-992A).

Data accessibility. All the uvCLAP, FLASH and RNA-seq data in this study have been deposited to the Gene Expression Omnibus database under the accession number GSE85164.
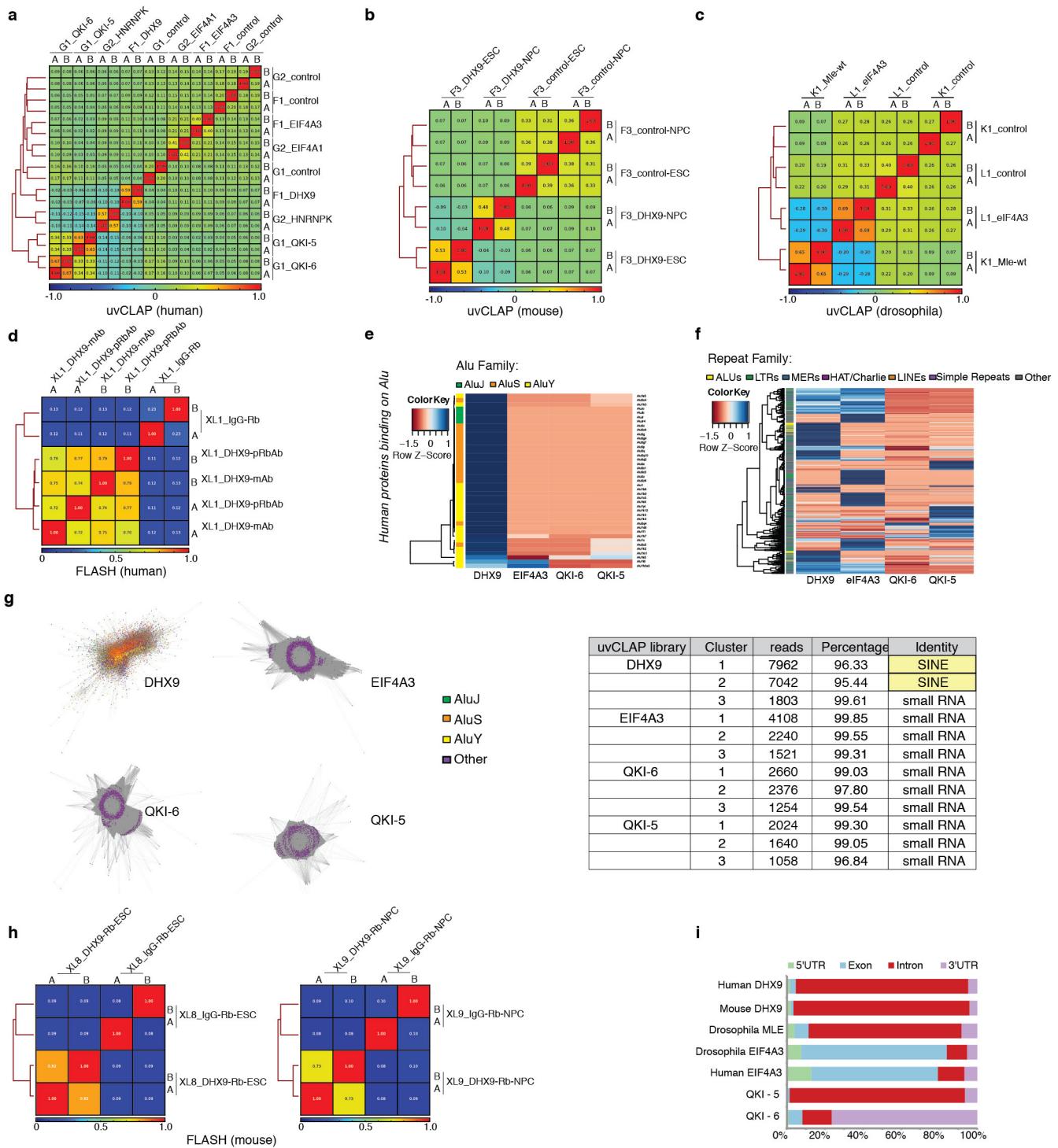
30. Shin, Y. J. *et al.* Nanobody-targeted E3-ubiquitin ligase complex degrades nuclear proteins. *Sci. Rep.* **5,** 14269 (2015).
31. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protocols* **8,** 2281–2308 (2013).
32. Chelmicki, T. *et al.* MOF-associated complexes ensure stem cell identity and Xist repression. *eLife* **3,** e02024 (2014).
33. Konig, J. *et al.* iCLIP--transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *J. Vis. Exp.* **50,** 2638 (2011).
34. Faircloth, B. C. & Glenn, T. C. Not all sequence tags are created equal: designing and validating sequence identification tags robust to indels. *PLoS One* **7,** e42543 (2012).
35. König, J. *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* **17,** 909–915 (2010).
36. Dodt, M., Roehr, J. T., Ahmed, R. & Dieterich, C. FLEXBAR—flexible barcode and adapter processing for next-generation sequencing platforms. *Biology (Basel)* **1,** 895–905 (2012).
37. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).
38. Ibrahim, M. M., Lacadie, S. A. & Ohler, U. JAMM: a peak finder for joint analysis of NGS replicates. *Bioinformatics* **31,** 48–55 (2015).
39. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44** (W1), W160–W165 (2016).
40. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* **44** (W1), W3–W10 (2016).
41. Holmqvist, E. *et al.* Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking *in vivo*. *EMBO J.* **35,** 991–1011 (2016).
42. Day, D. *et al.* Estimating enrichment of repetitive elements from high-throughput sequence data. *Genome Biol.* **11,** R69 (2010).
43. Novák, P., Neumann, P. & Macas, J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* **11,** 378 (2010).
44. Novák, P., Neumann, P., Pech, J., Steinhaisl, J. & Macas, J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* **29,** 792–793 (2013).
45. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14,** R36 (2013).
46. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30,** 923–930 (2014).
47. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
48. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* **22,** 2008–2017 (2012).
49. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
50. Steigemann, P. *et al.* Aurora B-mediated abscission checkpoint protects against tetraploidization. *Cell* **136,** 473–484 (2009).
51. Picardi, E. *et al.* REDIportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res.* **45,** D750–D757 (2017).

Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | uvCLAP/FLASH identifies DHX9 targets in humans and mice. a, The domain organization of *Drosophila*, human and mouse Mle/DHX9 is shown side by side with the other RNA-binding proteins, EIF4A3, QKI-5/6 used in uvCLAP, and the two other enzymatic-domain-containing dsRNA-binding proteins PKR and ADAR. Blue boxes show dsRNA-binding domains, green boxes show helicase-N and helicase-C domains, orange boxes show KH domains, pink box shows the Z-DNA-interacting domains of ADAR. Orange boxes with vertical lines show the RGG-repeats at the C-term of DHX9/Mle/RHA proteins that interact with single-stranded nucleic acids. **b**, The schematic representation of the two methods (uvCLAP and FLASH) developed for the identification of *in vivo* targets of RNA-binding proteins. **c**, The distance between peaks which do not directly overlap with an *Alu* repeat to the nearest *Alu* element for RNA-binding proteins EIF4A3, EIF4A1, HNRNPK, QKI-5 and QKI-6 is significantly further than similar peaks for DHX9. uvCLAP and FLASH data are shown separately which are put together in Fig. 1c. **d**, The fraction of peaks which are directly overlapping with an *Alu* element is depicted for DHX9 in comparison to peaks randomly placed in targeted regions within a sequencing library (shown as F1, G1, G2) and for EIF4A3, QKI-6, QKI-5, EIF4A1 and HNRNPK in uvCLAP and FLASH experiments (sequencing library XL1). **e**, The evolutionary divergence of primate and murine SINEs from the 7SL RNA. Scale bar, genetic distance (arbitrary units). **f**, The fraction of peaks which are directly overlapping with a SINE B1 repeat is depicted for DHX9 in comparison to peaks randomly placed in targeted regions. **g**, Fraction of uvCLAP and FLASH peaks in mouse embryonic stem (mES) cells and neural progenitor cells (NPCs) with a distance of at most 100 nucleotides from SINE B repeats is significantly enriched ($P < 2.2 \times 10^{-16}$; Fisher's exact test) for DHX9 in comparison to shuffled intron controls. **h**, The distances of DHX9 peaks in mES cells, which do not directly overlap with B1 repeats, to the nearest B1 repeats are shown. The distance is significantly smaller than the corresponding shuffled controls with median distances of 508 ($n = 33,686$) vs 1,312.5 ($n = 38,436$) nucleotides for uvCLAP and 514 ($n = 44,557$) vs 1,108.5 ($n = 51,272$) nucleotides for FLASH (all $P < 2.2 \times 10^{-16}$; one-tailed Mann–Whitney U-test). $n =$ number of uvCLAP/FLASH peaks.
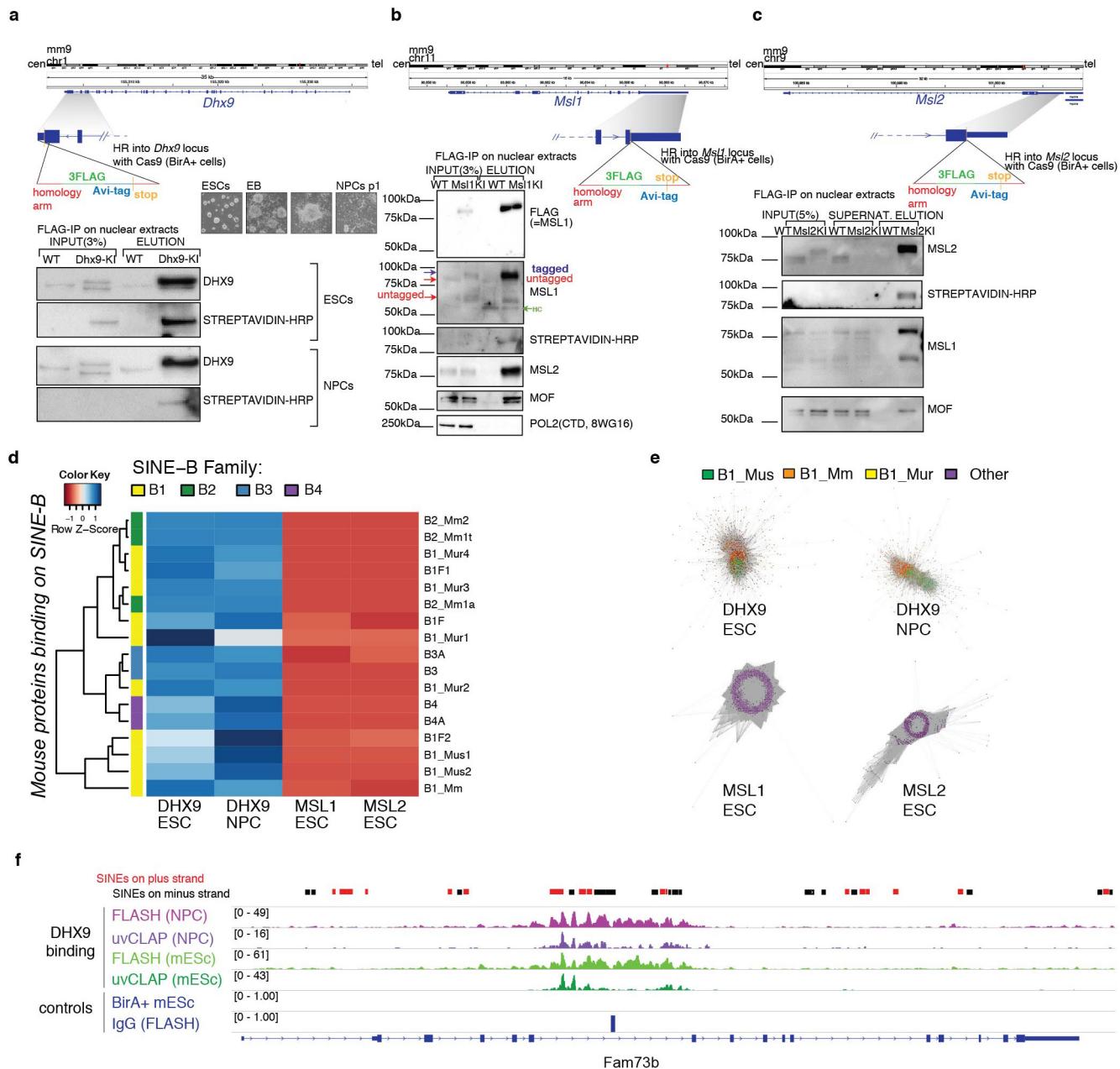
The uvCLAP library table:

uvCLAP library	Cluster	reads	Percentage	Identity
DHX9	1	7962	96.33	SINE
	2	7042	95.44	SINE
	3	1803	99.61	small RNA
EIF4A3	1	4108	99.85	small RNA
	2	2240	99.55	small RNA
	3	1521	99.31	small RNA
QKI-6	1	2660	99.03	small RNA
	2	2376	97.80	small RNA
	3	1254	99.54	small RNA
QKI-5	1	2024	99.30	small RNA
	2	1640	99.05	small RNA
	3	1058	96.84	small RNA

Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | Reproducibility of the replicates for each uvCLAP/FLASH library and enrichment of *Alu* repeat binding in human cells. a, Replicates (shown as A and B) in human uvCLAP sequencing libraries F1, G1 and G2 are compared to each other using Spearman correlation and similarity is represented as a heat map. Except for replicates the only other two profiles which correlate are of QKI isoforms. **b**, Replicates (shown as A and B) in mouse uvCLAP sequencing library F3, which is composed of DHX9 knock-in mES cells and NPCs, and BirA ligase control cell line, are compared to each other using Spearman correlation and similarity is represented as a heat map. Replicates of each experiment cluster together. **c**, Replicates (shown as A and B) in *Drosophila melanogaster* uvCLAP sequencing libraries K1 and L1, which are composed of Mle and Eif4a3 profiles, are compared to each other using Spearman correlation and similarity is represented as a heat map. Replicates of each experiment cluster together. **d**, Replicates (shown as A and B) in human FLASH sequencing library XL1 are compared to each other by using Spearman correlation. DHX9 library replicates as well as different antibody used in two independent library replicates cluster together, whereas IgG control is separated. **e**, Enrichments of *Alu*
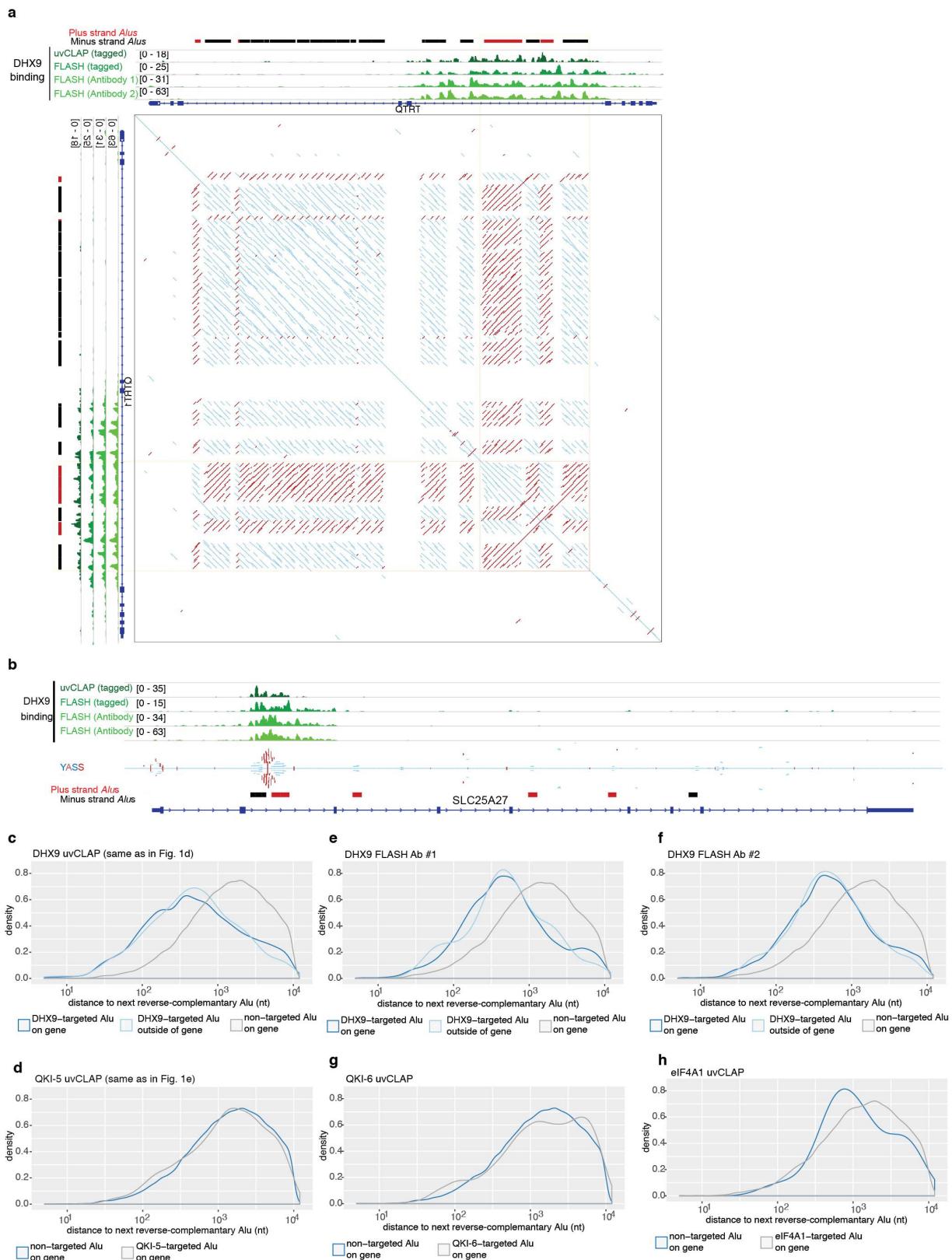
repeat subfamilies in uvCLAP data (see Methods). All subfamilies of *Alu* elements are highly enriched in the DHX9 uvCLAP experiment, compared to EIF4A3, QKI-5 and QKI-6. **f**, Enrichments of all human repeat families in uvCLAP data (see Methods). *Alu* elements are highly enriched in the DHX9 uvCLAP experiment, compared to EIF4A3, QKI-5 and QKI-6. **g**, Mapping-free clustering (see Methods) of repeated reads within uvCLAP libraries for human DHX9, EIF4A3, QKI-6 and QKI-5 reveals that only the top cluster of DHX9 library is composed of *Alu* elements. Right, a table view for the identity of the top three clusters (maximum reads) produced by graph based clustering of raw sequencing reads from different uvCLAP experiments. *Alu* elements bound by DHX9 form two distinct clusters (as depicted in the table) while the top three clusters in other samples (QKI-6, QKI-5 and eIF4A3) contain various small RNA families. **h**, Right, replicates (shown as A and B) in mouse FLASH sequencing library XL9 are compared to each other using Spearman correlation (see Supplementary Table 3 for mapping statistics and number of peaks). **i**, Binding frequencies of DHX9, EIF4A3 and QKI isoform crosslinking events on mRNA targets shown for UTRs, exons and introns.

Extended Data Figure 3 | Mapping for DHX9-bound RNAs onto repeats reveals an enrichment of SINE B binding in mouse cells.
a, Top, endogenous knock-in strategy is shown for the mouse *Dhx9* gene. Right, differentiation snapshots of the tagged mES cell line into NPCs. Bottom, the identity of the tagged protein and its biotinylation is shown by DHX9 and streptavidin blots from Flag immunoprecipitation (IP) carried in nuclear extracts both for mES cells and NPCs. **b**, Endogenous knock-in strategy is shown for the mouse *Msl1* gene. The integrity of the protein and its biotinylation is shown by MSL1 and streptavidin blots from Flag IP carried in nuclear extracts. The interaction ability of the tagged MSL1 protein with its known interaction partners (within MSL complex) is validated by this Flag IPs (performed on soluble nuclear extracts) by showing the co-IP for MOF (MYST1) and MSL2. POL2 (8WG16) is used as a loading control. **c**, Endogenous knock-in strategy is shown for the mouse *Msl2* gene. The integrity of the protein and its biotinylation is shown by MSL2 and streptavidin blots from Flag IP carried in nuclear
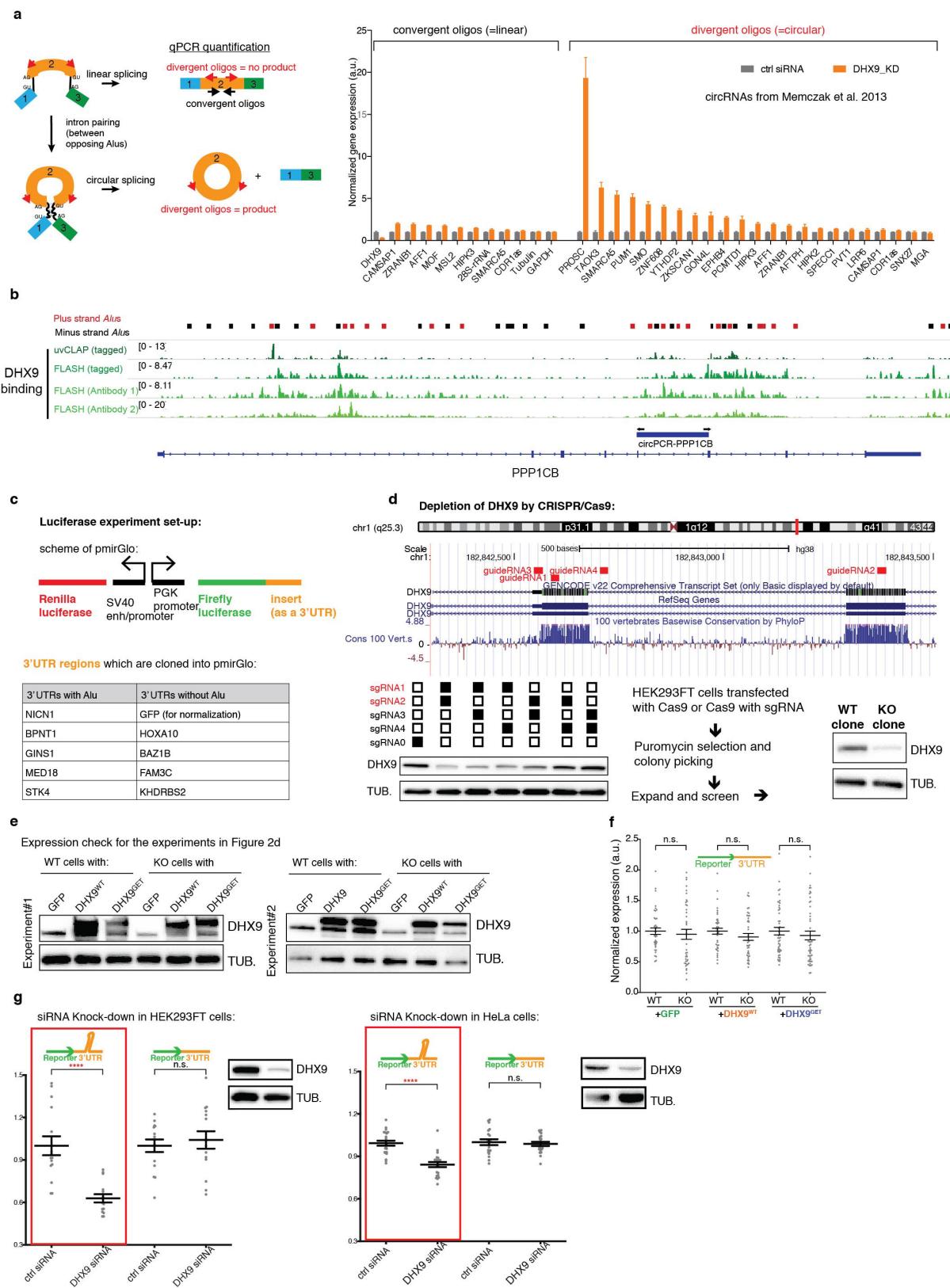
extracts. The interaction ability of the tagged MSL2 protein with its known interaction partners (within MSL complex) is validated by this Flag IP (performed on soluble nuclear extracts) by showing the co-IP for MOF (MYST1) and MSL1. **d**, Enrichments of SINE repeat subfamilies in uvCLAP data (see Methods). All SINE B subfamilies are highly enriched in DHX9 uvCLAP experiment, both in mES cells and NPCs, compared to MSL1 and MSL2. **e**, Mapping-free clustering of reads (see Methods) within uvCLAP libraries (mES cells and NPCs) reveals that only the top cluster of mouse DHX9 library is composed of SINE B repeats. **f**, Snapshots of uvCLAP and FLASH binding events (both in mES cells and NPCs) within *Fam73b* gene show that the crosslinking sites of DHX9 reside on SINE B repeats on opposing strands. SINEs on the plus strand are shown with red boxes, SINEs on the minus strand are shown with black boxes. Biological replicates were merged in these representations (also see Extended Data Fig. 2 for reproducibility between biological replicates).

Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | uvCLAP and FLASH reveal DHX9 binding on inverted-repeat *Alu* elements. a, A snapshot of DHX9 binding in the intron of the *QTRT1* gene. *Alu* elements are depicted with red (plus strand) or black (minus strand) boxes. Dot-plot generated by YASS (see Methods) showing that DHX9 binds to inverted repeats (structure-forming), shown by red lines, but not to direct repeats, shown by blue lines. **b**, A snapshot of DHX9 binding in the intron of the *SLC25A27* gene where there are *Alu* elements on opposite strands. YASS-generated dot-plots as in **a** (for simplicity only the diagonal of the YASS dot-plot is shown). **c**, Genome-wide analysis of uvCLAP DHX9 binding near paired *Alu* elements shows preferential targeting of *Alu* elements with nearby binding partners in both gene (dark blue) and outside gene (light blue) bound regions. **d**, Genome-wide analysis of uvCLAP QKI-5 binding near paired *Alu* elements (binding enriched in introns) shows no significant difference for closeness to nearest potential binding partner. **e**, Genome-wide analysis of FLASH DHX9 binding near paired *Alu* elements (Ab#1) shows a preferential targeting of *Alu* elements with nearby binding partners in both gene (dark blue) and outside gene (light blue) bound regions. **f**, Genome-wide analysis of FLASH DHX9 (Ab#2) binding near paired *Alu* elements shows a preferential targeting of *Alu* elements with nearby binding partners in both gene (dark blue) and outside gene (light blue) bound regions. **g**, Genome-wide analysis of uvCLAP QKI-6 binding near paired *Alu* elements shows no significant difference for closeness to nearest potential binding partner. **h**, Genome-wide analysis of uvCLAP EIF4A1 binding near paired *Alu* elements shows no significant difference for closeness to nearest potential binding partner.
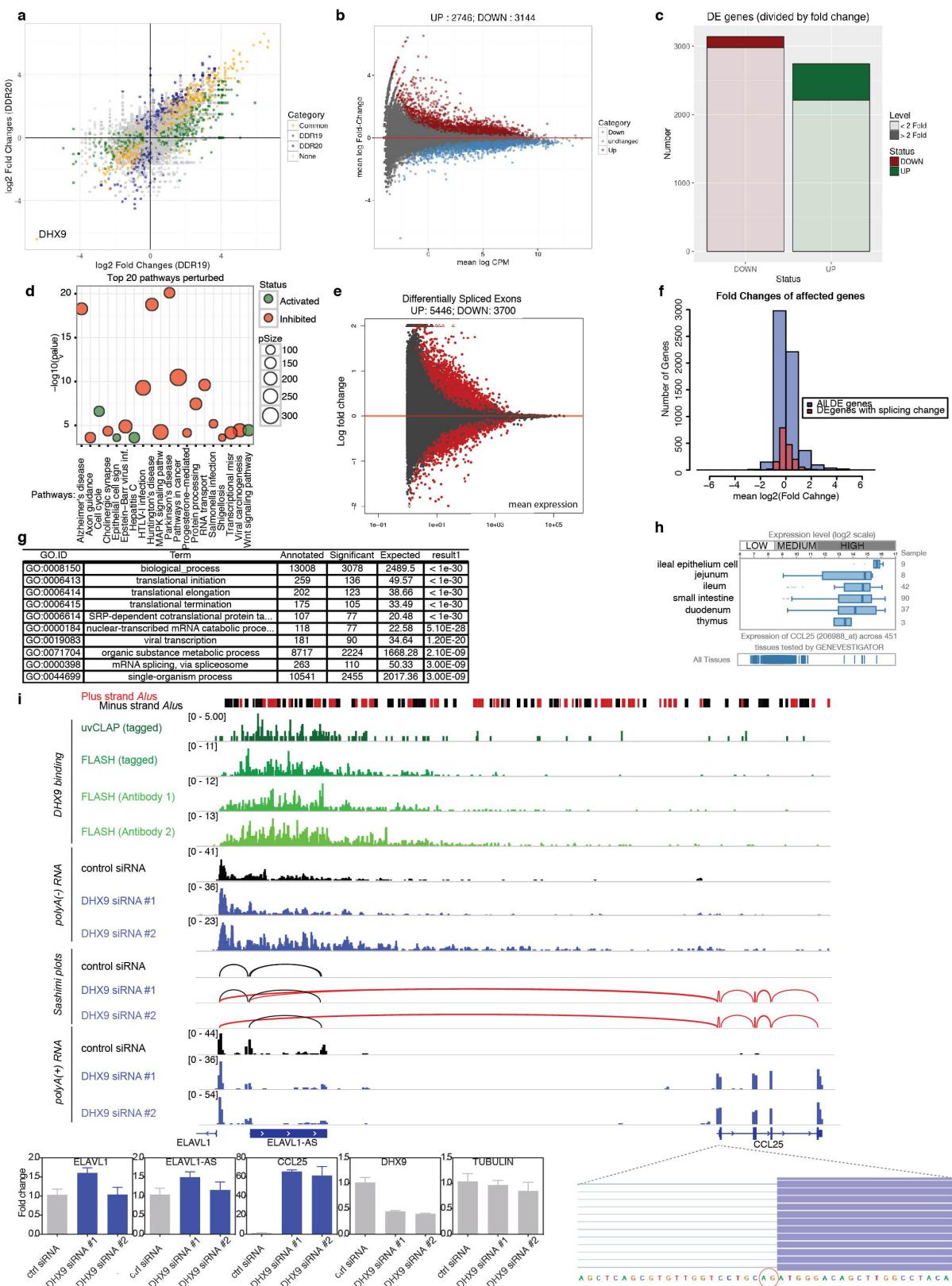
a

convergent oligos (=linear) — divergent oligos (=circular)

circRNAs from Memczak et al. 2013

b

DHX9 binding

PPP1CB

c

Luciferase experiment set-up:

scheme of pmirGlo:

Renilla luciferase — SV40 enh/promoter — PGK promoter — Firefly luciferase — insert (as a 3'UTR)

3'UTR regions which are cloned into pmirGlo:

3'UTRs with Alu	3'UTRs without Alu
NICN1	GFP (for normalization)
BPNT1	HOXA10
GINS1	BAZ1B
MED18	FAM3C
STK4	KHDRBS2

d Depletion of DHX9 by CRISPR/Cas9:

HEK293FT cells transfected with Cas9 or Cas9 with sgRNA
↓
Puromycin selection and colony picking
↓
Expand and screen →

e Expression check for the experiments in Figure 2d

f

g siRNA Knock-down in HEK293FT cells: siRNA Knock-down in HeLa cells:

Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | DHX9 depletion increases circular RNA formation and represses translation of reporters containing inverted-repeat *Alu* elements within the 3′ UTR. a, Left, the qPCR approach for detecting linear/circular RNAs with two sets of oligos indicated as divergent and convergent used for the detection of circular and linear RNAs, respectively. Right, quantitative real-time PCR analysis of previously reported circRNAs[23] in DHX9-siRNA-treated HEK293FT cells (error bars represent standard deviation between biological quadruplicates). **b**, Snapshot of a circRNA generated locus, PPP1CB with DHX9 crosslinking sites on inverted *Alu* repeats. Blue bar depicted as 'circPCR-PPP1CB' represents the qPCR scored region in (Fig. 2c). Biological replicates were merged in these representations (also see Extended Data Fig. 2 for reproducibility between biological replicates). **c**, Top, schematic drawing of pmirGlo insert cloning is shown. Bottom, 3′ UTRs used as inserts in luciferase assays. **d**, Top, genomic position of tested guide RNAs in CRISPR/Cas9 depletion of DHX9 is shown
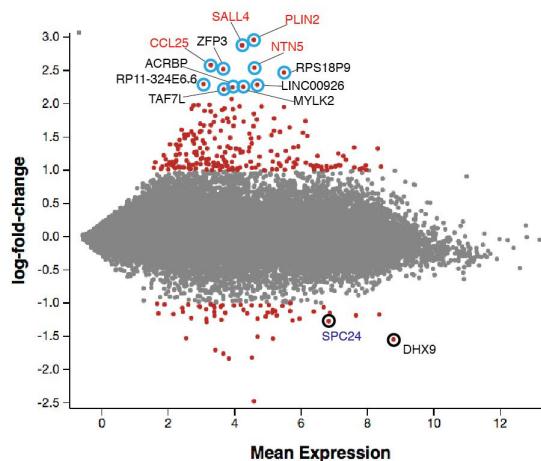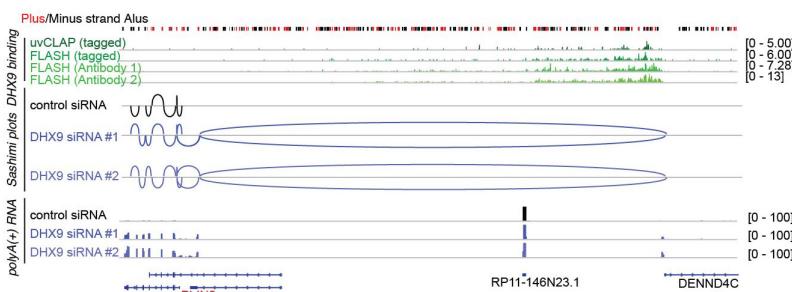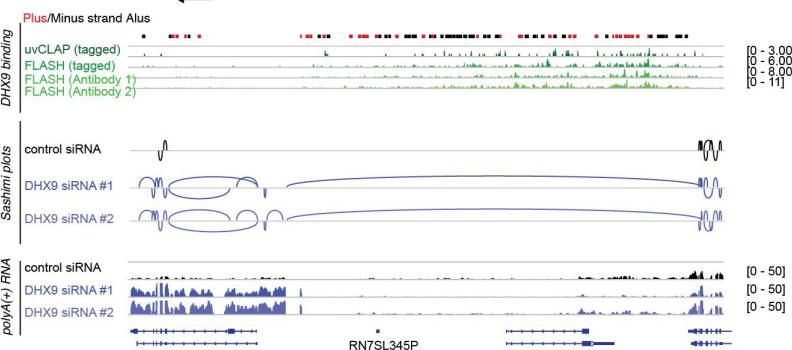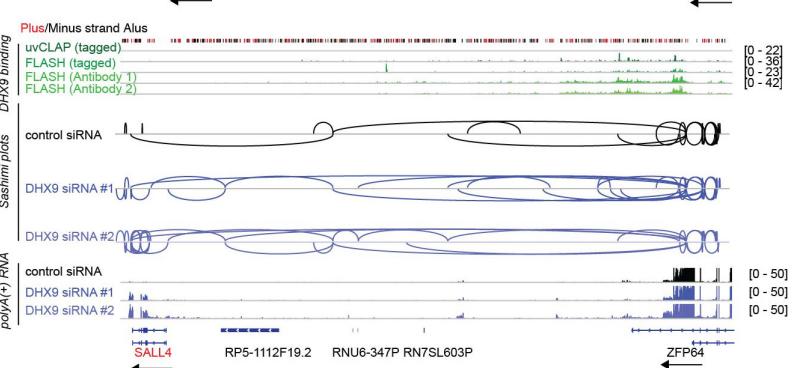
(see Supplementary Table 2 for guide RNA sequences). Bottom left, efficiency of guide RNA pairs is shown by DHX9 and tubulin western blots. Bottom right, description for the making of the constitutive DHX9-depleted clone. **e**, Expression check by western blot analysis of rescue cells used on two independent (different days) experiments. **f**, Luciferase assays show that DHX9-depletion does not alter the expression of 3′ UTR elements without inverted-repeat *Alu* elements. **g**, Luciferase assay results are shown carried in HEK293FT and HeLa cells upon siRNA knockdown of DHX9. Knockdown efficiency is validated by western blot analysis for DHX9 and tubulin. Similar to what is shown in Fig. 2d reporter expression from the constructs with an inverted-repeat *Alu* element in 3′ UTRs are affected upon DHX9 depletion. Error bars represent standard deviation of a total of 20 data points that come from one experiment carried out with biological quadruplicates for each cloned 3′ UTR insert (5 with and 5 without inverted-repeat *Alu* elements).

Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | DHX9 depletion leads to changes in gene expression and exon usage. a, Reproducibility of gene-expression changes between cells treated with siRNA 1 (green) and siRNA 2 (blue) (both show that DHX9 is the most severely downregulated gene in these datasets). Genes with a significant change in their expression levels in both experiments are shown with yellow points. **b**, MA plot of genes that show a reproducible change in both RNAi experiments (yellow dots in **a**). **c**, The extent of gene expression changes in the reproducibly misregulated genes in **b**. **d**, Pathway perturbation analysis using the data in **b**. **e**, Differentially spliced exons that show a reproducible change in DHX9-depletion experiments (siRNA 1 and siRNA 2). **f**, Differentially expressed genes (DEgenes) with respect to their detected splicing changes shows 55% of the mis-spliced genes are downregulated. **g**, GO term analysis using data in **e**. **h**, Tissue expression database (Genevestigator) results for *CCL25* gene show its expression is specific to the intestine and thymus tissues. CCL25 is expressed at low levels every other tissue (in total 451 tissues). **i**, DHX9
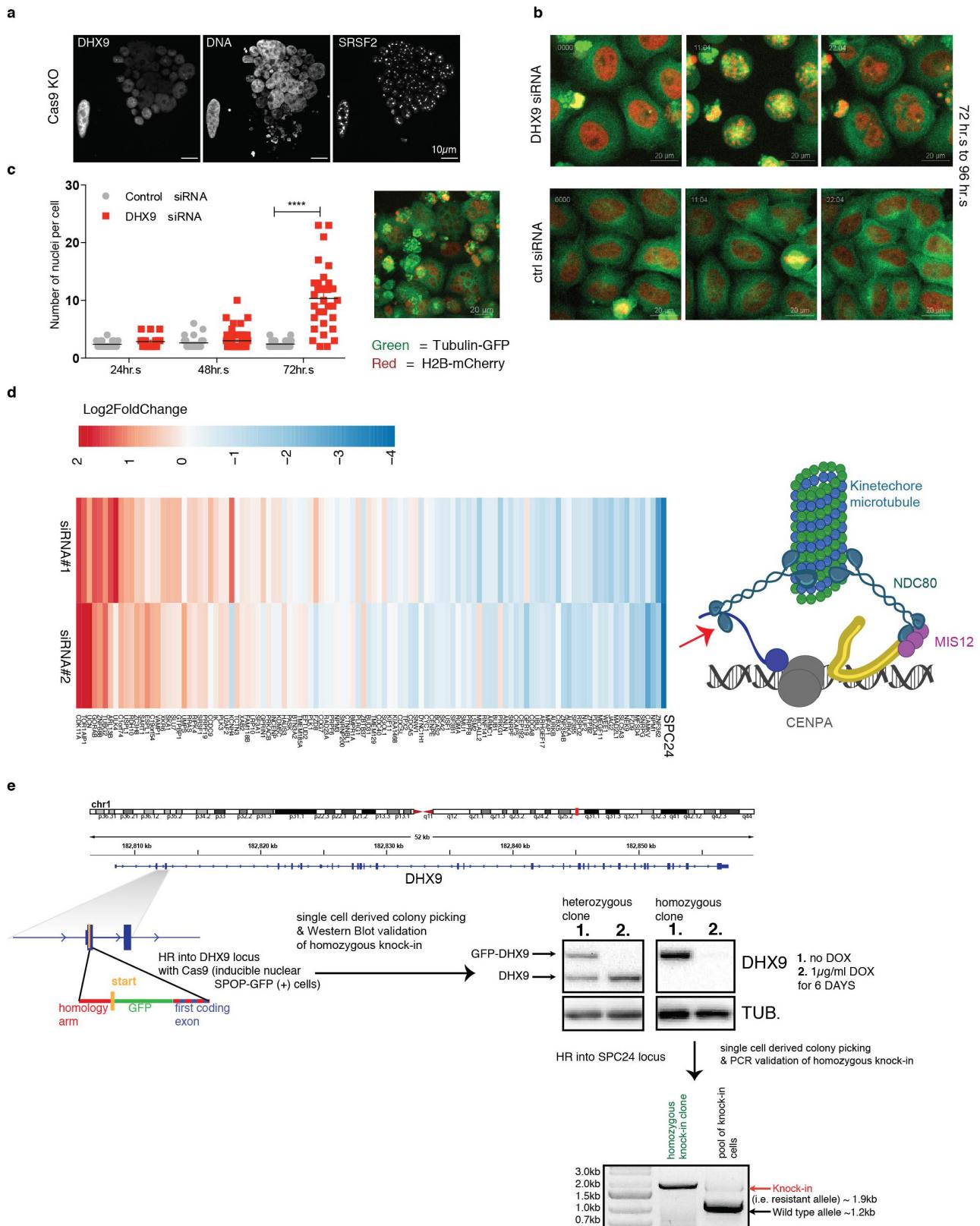
binding in the *ELAVL1–CCL25* locus is shown together with the poly(A)$^-$ and poly(A)$^+$ RNA-seq data. Sashimi plots generated from control and DHX9-siRNA-treated samples show a new exon–exon junction from *ELAVL1^{AS}* to *CCL25* (depicted with red lines). For clarity, only plus strand data are shown. Biological replicates were merged in these representations (also see Extended Data Fig. 2 for reproducibility between biological replicates). Poly(A)$^+$ RNA-seq data show that an anti-sense transcript in the opposite direction to *ELAVL1* is now connected via a cryptic splice acceptor site to the first exon of *CCL25* upon DHX9 depletion (two independent siRNAs, four biological replicates each). Bottom, qPCR validation of the CCL25 upregulation upon RNA processing defects in DHX9 knockdown samples. We observe that CCL25 is ~60-fold upregulated in DHX9-depleted cells, while the expression of neither *ELAVL1* nor its accompanying anti-sense transcript change significantly. The cryptic splice acceptor site is enlarged on the bottom right.

Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Many upregulated genes in DHX9 knockdown show RNA processing defects in their gene locus. a, Genes that were more than twofold up- or downregulated are marked with red dots. Top 11 upregulated genes, which show signs of transcription bleed-through from upstream genes are highlighted with blue circles. *CCL25*, *SALL4*, *PLIN2* and *NTN5* are further highlighted. **b**, DHX9 binding in the *PLIN2–RP11–146N23.1* gene locus on the nascent RNA originating from the antisense transcription of the *DENND4C* promoter and splicing defect upon DHX9 knockdown. Sashimi plots depicting the exon junctions of these genes are shown in blue for the DHX9 knockdown, whereas the control sample (shown with black lines) does not display such a joining event (threshold is set to five reads). Biological replicates were merged in these representations (also see Extended Data Fig. 2 for reproducibility between biological replicates). For clarity, only minus strand data are shown. **c**, DHX9 binding in the *NTN5–MAMSTR* gene locus on the nascent (bleed-through) RNA and splicing defect upon DHX9 knockdown. Sashimi plots depicting the exon junctions of these genes are shown in blue for the DHX9 knockdown, whereas the control sample (shown with black lines) do not display such a joining event (threshold is set to one read). Biological replicates were merged in these representations (also see Extended Data Fig. 2 for reproducibility between biological replicates). For clarity, only minus strand data are shown. **d**, DHX9 binding in the *SALL4–ZFP64* gene locus on the nascent (bleed-through) RNA and splicing defect upon DHX9 knockdown. Sashimi plots depicting the exon junctions of these genes are shown in blue for the DHX9 knockdown whereas the control sample (shown with black lines) displays a lower frequency of such joining events (threshold is set to 1 reads). Biological replicates were merged in these representations (also see Extended Data Fig. 2 for reproducibility between biological replicates). For clarity, only minus strand data are shown.
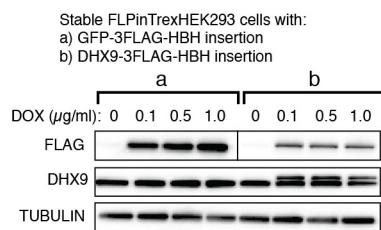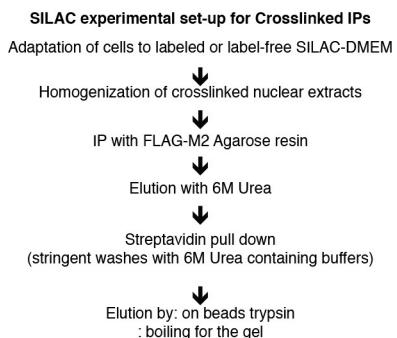
a

b

c

Green = Tubulin-GFP
Red = H2B-mCherry

d

Log2FoldChange

e

Extended Data Figure 8 | See next page for caption.
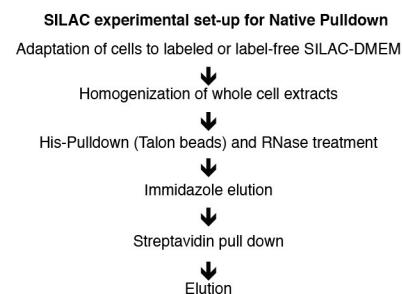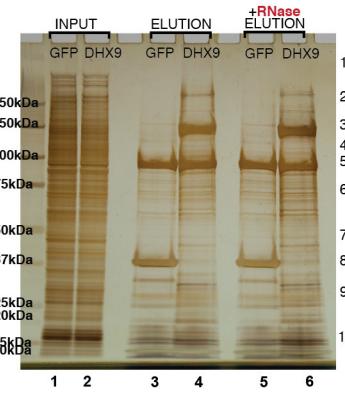
Extended Data Figure 8 | DHX9 depletion leads to a catastrophic disintegration of the nucleus. a, CRISPR–Cas9 DHX9-knockout cell (on the right side of the image) displays a multinucleated cell phenotype in comparison to a DHX9-expressing cell (on the left side of the image). Scale bar, 10 μm. **b,** Three still images from Supplementary Video 1 (control) and Supplementary Video 3 (knockdown) representing the start, middle and end point of 22-h-long live imaging for control or DHX9-siRNA-treated HeLa H2B–mCherry, tubulin–GFP cells. **c,** Quantification of the number of nuclei at the end point of live cell imaging (∼24 h imaging duration). Time points on the graph represent the start point of the imaging. All the imaged fields are taken into consideration and only the cells with more than one nucleus are included in the analysis. Statistical analysis performed with Kruskal–Wallis test shows that only the 72–96 h time point of DHX9-knockdown cells significantly differs from the rest with a $P < 0.0001$ (Kruskal–Wallis). Also see Supplementary Videos 2 and 7 for full field views of imaging and Supplementary Videos 5 and 6 for a second siRNA knockdown of DHX9. Right, multinucleated cells at the end point of imaging (at 96 h of DHX9 knockdown) in a different field from Supplementary Video 7. **d,** Left, expression levels for the Mitocheck Consortium Grape phenotype category genes (117 are expressed in the RNA-seq out of the 153 in total, mean counts per million > 0). Right, schematic view of the Ndc80 complex, red arrow points at the SPC24 protein. **e,** Left, experimental design of the homologous recombination of the GFP tag at the DHX9 locus is shown. Single-cell-derived heterozygously or homozygously GFP-tagged clones successfully deplete the GFP–DHX9 protein upon doxycycline induction of SPOP–GFP. Right, PCR-based screening of the knock-in allele is shown for pool of cells before the colony picking and the homozygous knock-in *Alu*-bypass SPC24 clone after isolation. Knock-in causes a size shift of the PCR product (from 1.2 to 1.9 kb). Correct insertion of the repair construct into the endogenous locus is further validated by Sanger sequencing (not shown).

a

Stable FLPinTrexHEK293 cells with:
a) GFP-3FLAG-HBH insertion
b) DHX9-3FLAG-HBH insertion

	a				b		
DOX (µg/ml)	0	0.1	0.5	1.0	0	0.1	0.5	1.0

FLAG
DHX9
TUBULIN

b

SILAC experimental set-up for Crosslinked IPs

Adaptation of cells to labeled or label-free SILAC-DMEM
↓
Homogenization of crosslinked nuclear extracts
↓
IP with FLAG-M2 Agarose resin
↓
Elution with 6M Urea
↓
Streptavidin pull down
(stringent washes with 6M Urea containing buffers)
↓
Elution by: on beads trypsin
: boiling for the gel

c

On-beads trypsin digest forward and reverse experiments

Gene names	Ratio H/L DHX9_E1_fw	Ratio H/L DHX9_E2_rev	INVERSE Ratio DHX9_E2_rev (L/H)	Average enrichment
DHX9	23.414	0.014	69.247	46.331
ADAR	5.847	0.076	13.106	9.476
DHX30	NaN	0.012	83.333	83.333
ZFR	9.158	0.043	23.017	16.087
ILF3	3.784	0.108	9.270	6.527
MATR3	7.751	0.038	26.501	17.126
SFPQ	10.324	0.031	32.687	21.506
HNRNPM	11.153	0.030	33.735	22.444
NONO	10.394	0.038	26.459	18.427
PSPC1	NaN	0.045	22.300	22.300
EIF4A3	4.165	0.121	8.267	6.216
HNRNPA1;HNRNPA1L2	9.413	0.034	29.036	19.224
ELAVL1	7.118	0.072	13.870	10.494
RALY	5.287	0.066	15.202	10.245
HNRNPC	11.750	0.026	38.814	25.282

d

SILAC experimental set-up for Native Pulldown

Adaptation of cells to labeled or label-free SILAC-DMEM
↓
Homogenization of whole cell extracts
↓
His-Pulldown (Talon beads) and RNase treatment
↓
Immidazole elution
↓
Streptavidin pull down
↓
Elution

e

f

Gene names	INVERSE Ratio DHX9_E2_rev (L/H) no RNase	INVERSE Ratio DHX9_E2_rev (L/H) with RNase	Fold change upon RNase
DHX9	235.593	9.293	25.352
ADAR	41.596	14.547	2.859
DHX30	156.397	13.822	11.315
ZFR	Not detected	Not detected	N.A.
ILF3	28.891	12.895	2.240
MATR3	16.906	4.504	3.754
SFPQ	1.943	1.546	1.257
HNRNPM	35.845	5.828	6.150
NONO	1.761	1.260	1.398
PSPC1	2.078	Not detected	N.A.
EIF4A3	29.526	Not detected	N.A.
HNRNPA1;HNRNPA1L2	Not detected	Not detected	N.A.
ELAVL1	46.320	2.664	17.387
RALY	70.378	3.977	17.697
HNRNPC	177.450	2.451	72.385

mRNA stability related NF45 interacting Paraspeckles associated

g

h

Interferon alpha (IFNa) induced ADAR-p150 and DHX9
Co-IP experiments in mouse (3T3) cells:

1. no treatment
2. 1000U/ml IFNa treatment

→ Whole-cell extract (WCE)
→ Nuclear extract (Nucl.)
→ Cytoplasmic extract (Cyto.)

→ Validation of Induction and Fractionation (i)
Endogenous DHX-IP with RNase treated lysates and
ADARp150 Co-IP (j,k,l)

i

j

Whole Cell Extract

k (same as in Figure 4c)

Nuclear Extract

l

Cytoplasmic Extract
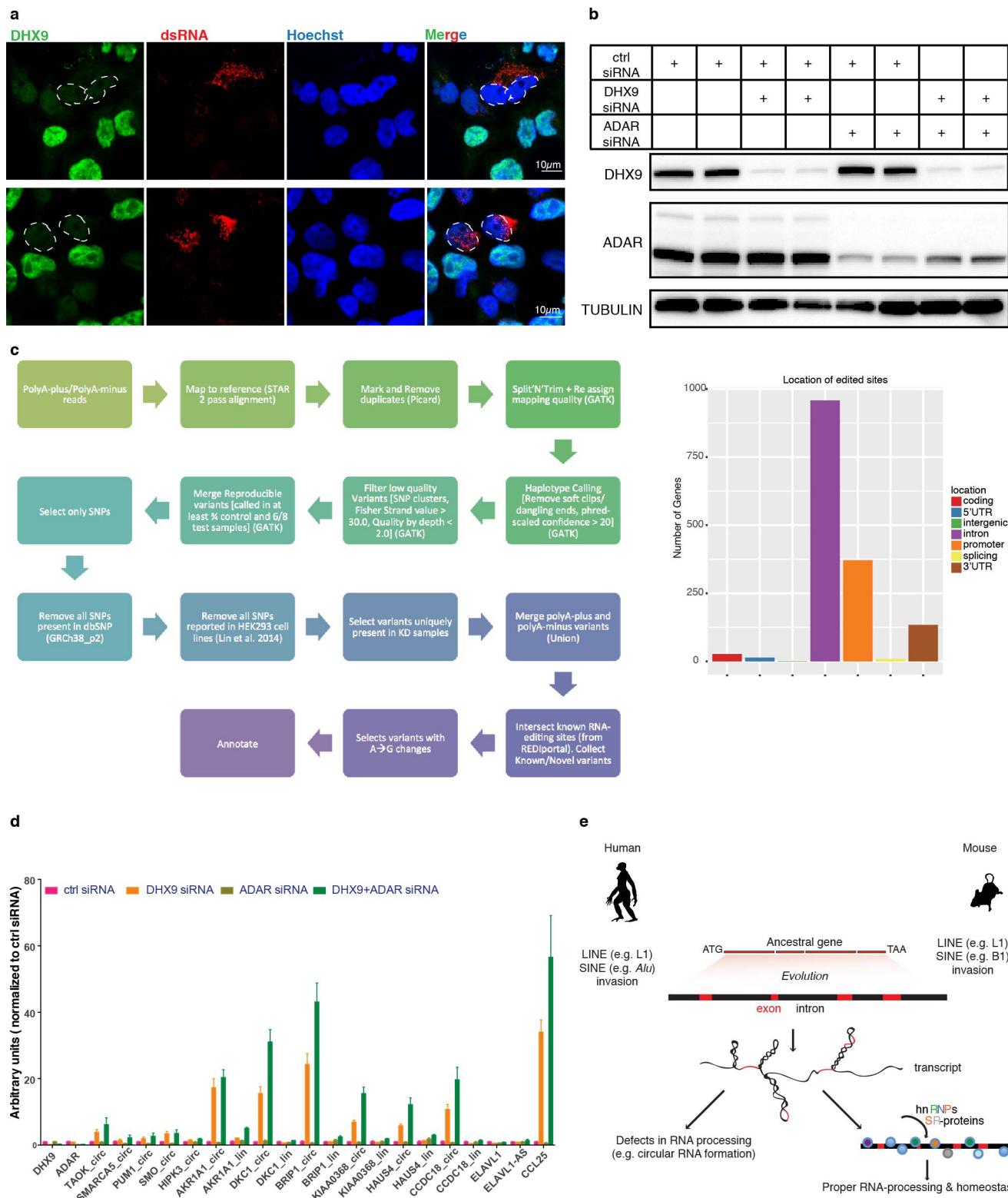
Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Tandem affinity purification of DHX9 from crosslinked nuclear extracts or native whole-cell extracts identifies a set of RNA-binding proteins that interact with DHX9. **a**, Induction test for the stable FLPinTrex HEK293 cell line with a single-copy C-terminal 3×Flag–6×His–biotin–6×His-tagged GFP or DHX9 protein. **b**, Experimental set-up for the SILAC quantified formaldehyde crosslinked DHX9 interactome (either loaded on gel or eluted from beads by trypsinization) obtained by tandem affinity purification. **c**, Fold enrichment values of on bead digested proteins from forward and reverse SILAC experiments (forward: DHX9 cells are grown in heavy isotope labelled SILAC medium and GFP cells are grown in unlabelled medium; reverse: DHX9 cells are grown in unlabelled medium and GFP cells are grown in heavy isotope labelled SILAC medium) for those proteins highlighted in Fig. 4a. **d**, Experimental set-up for the affinity tag pull-downs employing native whole-cell extracts that were used for in-gel digestion LC–MS. **e**, Silver-stain gel of native pull-down experiments, where lanes 1 and 2 show the input from GFP and DHX9 cell lines (DHX9 cells are grown in unlabelled medium and GFP cells are grown in heavy

isotope labelled SILAC medium), lanes 3 and 4 are the final eluates of His-tag pull-down followed by streptavidin pull-downs from GFP and DHX9 cell lines, and lanes 5 and 6 are RNaseA-treated final eluates. The gel is cut into 10 slices as shown and both GFP and DHX9 lanes are combined (3 and 4 together; 5 and 6 together) prior to mass-spectrometry analysis. **f**, Fold enrichment values in the native purifications for the proteins highlighted in Fig. 4a, showing a depletion for most of these proteins upon RNase digestion. **g**, Validation of DHX9 interaction with ADAR is shown in a stable cell line for ADAR that can express both p110 and p150 isoforms. The doublet is detected with a blot for Flag after DHX9 IP in this stable cell line. POL2 and CDK9 blots serve as loading controls. **h**, The description of the experimental set-up for IFNα induction and DHX9–ADAR co-IP in different fractions of the mouse 3T3 cell line. **i**, Validation of fractionation with tubulin (lanes 3, 4) and RNA–POL2 blots (lanes 2, 5) and IFNα induction of ADAR(p150) (lanes 5 and 6). **j**–**l**, Mouse DHX9 interacts with ADAR(p150) in the whole cell (**j**, lanes 3 and 6) and nuclear extracts (**k**, lanes 3 and 6) but not in the cytoplasmic extracts (**l**, lanes 3 and 6).

Extended Data Figure 10 | See next page for caption.

Extended Data Figure 10 | DHX9 and ADAR cooperatively work on large RNA secondary structures. a, Anti-DHX9 and J2 (dsRNA) antibody staining are shown for a mixed population of cells treated with control siRNA (DHX9 staining present, green) and DHX9 siRNA (DHX9 staining absent, for example, marked nuclei). DHX9-depleted cells accumulate dsRNA. **b**, The efficiency of knockdown in replicate experiments for J2 pull-down experiment displayed in Fig. 4e are shown by western blot. **c**, Left, analysis pipeline for detecting differential editing between control and knockdown RNA-seq libraries (both poly(A)$^+$ and poly(A)$^-$ considered). Right, distribution of differential A-to-I editing sites detected in DHX9 knockdown is shown. After filtering for known SNPs from dbSNP and HEK293 genome, 1,244 genes with 1,807 potential A-to-I RNA editing changes in DHX9-siRNA-treated samples were detected which were absent in controls. 48% of these sites have been reported previously[51]. 77% of genes showing editing changes have these changes within introns. Additionally, we observed that 21% of genes with intronic editing also showed splicing defects and 28% of genes with overall editing showed misregulation (differentially expressed) upon DHX9 knockdown. These genes are a small fraction of all the genes with splicing defects (4.5%) or expression changes (6%) observed in DHX9-depleted cells. **d**, Effect of DHX9, ADAR or double knockdown of DHX9 and ADAR on circular RNA generation and CCL25 upregulation is shown. ADAR depletion alone does not have an appreciable impact on circular RNA generation, whereas the effect of DHX9 depletion is augmented when it is combined with ADAR depletion. **e**, A summary model. *Alu* elements are the most abundant transposable elements making up >10% of our own genomes. These elements are potentially harmful to their host and are depleted entirely from 5′ UTRs of genes and developmentally important loci, such as the *HOXA–D* clusters but are found in abundance in intronic and intergenic regions. We show that DHX9, a highly conserved, very abundant nuclear RNA helicase, directly interacts with *Alu* elements *in vivo* and suppresses circular RNA formation, which is probably a symptom of *Alu*-mediated splicing defects. We propose that our cells have developed a dependency on DHX9 to remove strong secondary structures, such as the ones originating from *Alu* insertions in the transcribed parts of our genome. We also show that in mice, DHX9 interacts with SINE B elements that are the murine equivalent of *Alu* elements, underscoring the flexibility of DHX9-mediated control of retrotransposon toxicity throughout evolution.