



# The regulation and functions of DNA and RNA G-quadruplexes

Dhaval Varshney<sup>1,4</sup>, Jochen Spiegel<sup>1,4</sup>, Katherine Zyner<sup>1,4</sup>, David Tannahill<sup>1</sup> and Shankar Balasubramanian<sup>1,2,3</sup>✉

**Abstract** | DNA and RNA can adopt various secondary structures. Four-stranded G-quadruplex (G4) structures form through self-recognition of guanines into stacked tetrads, and considerable biophysical and structural evidence exists for G4 formation in vitro. Computational studies and sequencing methods have revealed the prevalence of G4 sequence motifs at gene regulatory regions in various genomes, including in humans. Experiments using chemical, molecular and cell biology methods have demonstrated that G4s exist in chromatin DNA and in RNA, and have linked G4 formation with key biological processes ranging from transcription and translation to genome instability and cancer. In this Review, we first discuss the identification of G4s and evidence for their formation in cells using chemical biology, imaging and genomic technologies. We then discuss possible functions of DNA G4s and their interacting proteins, particularly in transcription, telomere biology and genome instability. Roles of RNA G4s in RNA biology, especially in translation, are also discussed. Furthermore, we consider the emerging relationships of G4s with chromatin and with RNA modifications. Finally, we discuss the connection between G4 formation and synthetic lethality in cancer cells, and recent progress towards considering G4s as therapeutic targets in human diseases.

Nucleic acids have considerable potential to fold into three-dimensional, ‘secondary’ structures. This can happen through the formation of non-Watson–Crick hydrogen bonds between nucleobases. Early observations on the self-assembly of guanylic acid<sup>1</sup> led to the elucidation of the guanine tetrad-forming sequence motif<sup>2</sup> (FIG. 1a), in which guanines are mutually bonded by Hoogsteen hydrogen base-pairing to form a planar array that is further stabilized by interactions between positively charged ions and the O-6 lone-pair electrons of each guanine (FIG. 1b,c). Initial evidence for the assembly of four-stranded G-quadruplex (G4) structures from natural sequences was provided by the formation in vitro of higher-order secondary structures from oligonucleotides resembling G-rich sequences from immunoglobulin switch regions<sup>3</sup>. Biophysical and structural biology methods subsequently provided substantial physical evidence for the formation of intermolecular and intramolecular G4s from DNA and RNA in vitro, including a framework for recognizing sequences likely to fold into G4s.

Although G4s are related to each other in primary sequence, they in fact comprise a diverse family of structures that can fold into various topologies, which are dictated by the pattern of strand polarities and also the orientation of interconnecting loops<sup>4</sup> (FIG. 1d).

The extent to which distinct topologies can influence G4 formation and function in cells is unknown. There has been a recent surge in research activity directed towards understanding G4 formation in living cells and organisms. Considerable attention has focused on the detection and occurrence of G4 structures in genomes and in RNA with a view to elucidating how these elements might regulate key biological processes, such as transcription, telomere homeostasis and translation. Identifying specific proteins that directly interact with G4 structures and elucidating their influence on such processes is an important step towards increasing our understanding of G4 biology. Detailed structural investigations into the mechanisms of G4 unfolding by helicases are providing insights into the control of G4 folding at the biochemical level, although further work is needed to fully understand the regulation of G4 formation in cells. The fact that G4s are linked with DNA damage and genome instability in addition to key cancer-associated genes has prompted investigations into possible roles of G4s in cancer biology and an evaluation of small-molecule G4 ligands as potential therapeutic agents.

In this Review, we discuss the evidence for G4 formation in DNA and RNA in biological systems, factors that regulate G4 formation and biological processes

<sup>1</sup>Cancer Research UK Cambridge Institute, Li Ka Shing Centre, Cambridge, UK.

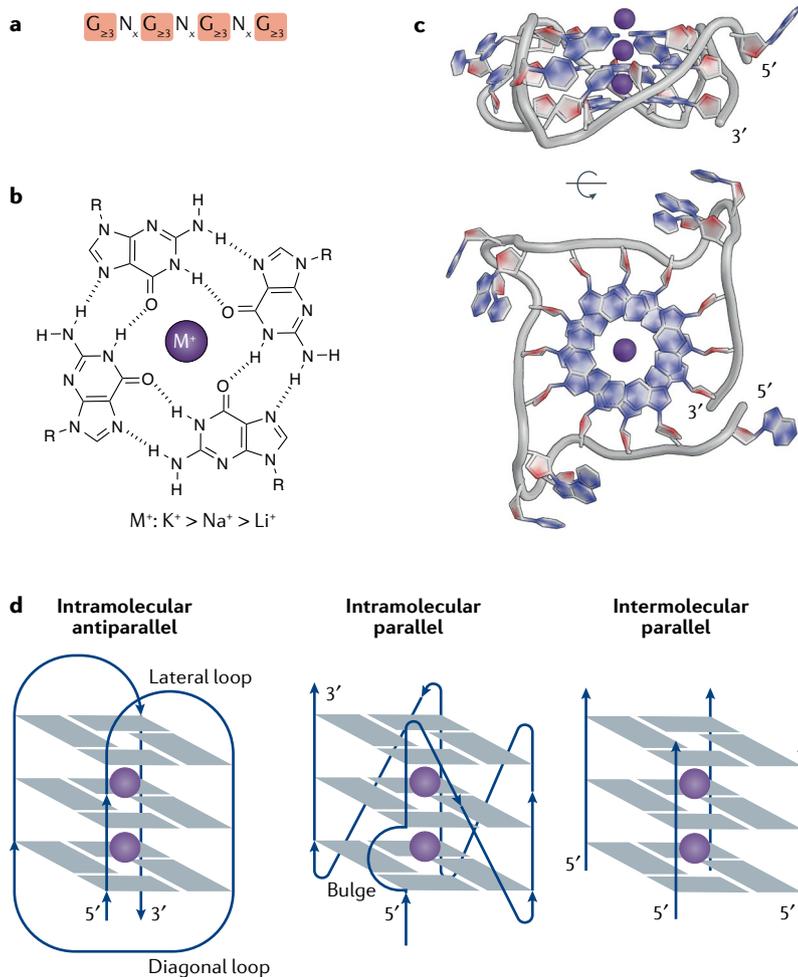
<sup>2</sup>Department of Chemistry, University of Cambridge, Cambridge, UK.

<sup>3</sup>School of Clinical Medicine, University of Cambridge, Cambridge, UK.

<sup>4</sup>These authors contributed equally: Dhaval Varshney, Jochen Spiegel, Katherine Zyner.

✉e-mail: sb10031@cam.ac.uk

<https://doi.org/10.1038/s41580-020-0236-x>



**Fig. 1 | The structure and topologies of G-quadruplexes.** **a** | The G-quadruplex (G4) consensus sequence. *x* denotes the number of nucleotides in the loops (see part **d**). **b** | A guanine tetrad is stabilized by Hoogsteen base-pairing and by a central cation ( $M^+$ ), with a preference for monovalent cations in the order of potassium ( $K^+$ ) > sodium ( $Na^+$ ) > lithium ( $Li^+$ ). **c** | X-ray crystal structure of an intramolecular, parallel G4 from a human telomere sequence (PDB: 1KF1)<sup>214</sup>. **d** | Schematic representation of some G4 topologies.

**Circular dichroism**

A spectroscopic technique to investigate structure based on the interaction of plane-polarized light with a structurally asymmetric molecule.

**Bayesian predictions**

Statistical methods to infer probabilities for a hypothesis, which can be updated when new information becomes available.

**G-fraction**

The proportion of G bases in a sequence, that is, G-richness.

**G-skew**

The under-representation or over-representation of G bases in a sequence.

that are influenced by G4s. We also discuss important links between G4s and cancer, and the progress made towards using G4s as therapeutic targets. Other reviews provide extensive details on G4 prediction<sup>5</sup>, biophysics and structure<sup>4,6</sup>, and roles in DNA replication<sup>7</sup>, human disease<sup>8</sup> and therapeutic possibilities<sup>9</sup>.

**Identification of G4s**

Biophysical studies using oligonucleotides were the first to establish that many DNA and RNA sequences featuring G-tracts separated by other bases (loops) can fold into G4 structures. Rules for predicting G4 structure formation emerged on the basis of data from circular dichroism, ultraviolet melting and NMR spectroscopy studies on different G4-forming oligonucleotides<sup>4,10</sup>. G4s have been identified as cellular features through a combination of computational sequence analyses and experiments that detect G4s in cellular genomes and in purified nucleic acids using chemical and molecular biology and imaging methods.

**Computational identification of G4s**

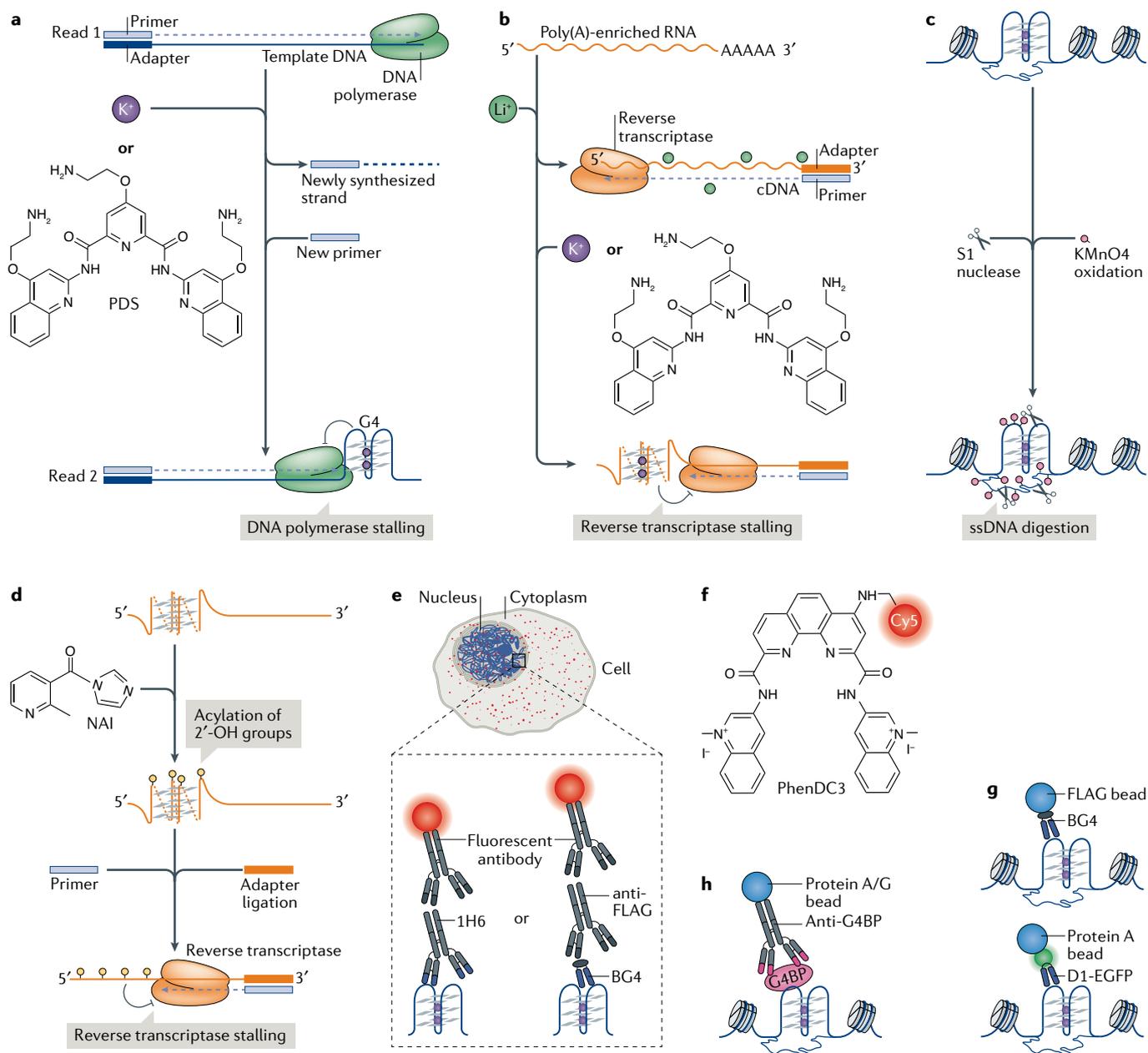
The use of early algorithms to search for the relatively simple  $G_{3-5}N_{1-7}G_{3-5}N_{1-7}G_{3-5}N_{1-7}G_{3-5}$  consensus sequence suggested that the human genome may have over 300,000 sequences with the potential to form G4s<sup>11,12</sup>. Such computational tools have helped to identify potential G4s associated with key genes and reveal enrichment of G4-forming sequences in genomic regions associated with gene regulation, specific cellular functions and disease states<sup>13,14</sup>. The early search algorithms were unable to account for structural variants, such as G4s with longer loops, bulges or mismatches, for two-tetrad G4s or for the importance of flanking sequences. More recent computational tools accommodate some of these factors<sup>15</sup>, use Bayesian predictions<sup>16</sup>, account for flanking sequence effects based on G-fraction and G-skew<sup>17</sup> or consider possibilities of higher-order assemblies<sup>18</sup>. Although machine learning approaches have helped to identify G4s that are likely to form in genomic contexts<sup>19</sup> and those likely to fold in RNA<sup>20</sup>, computational methods that account for chromatin contexts and protein binding are yet to be developed.

**DNA and RNA chain-extension stalling**

Experimental approaches have been developed to detect G4 structures and complement computational prediction. G4s in DNA or RNA can stall a DNA polymerase<sup>21,22</sup> or a reverse transcriptase<sup>23</sup>, respectively. Comparison of polymerase pause sites in G4-stabilizing conditions (for example, in the presence of potassium ions ( $K^+$ ) and/or a stabilizing ligand) and in conditions that do not stabilize G4s (for example, in the presence of lithium ions ( $Li^+$ )) enables the detection of the 5' end of G4s in vitro. This method was initially applied using sequence analysis on a polyacrylamide gel and, more recently, adapted into a genome-wide DNA polymerase-stop assay followed by high-throughput sequencing (G4-seq; FIG. 2a). G4-seq identified more than 700,000 DNA G4 sites in the human genome<sup>24</sup>, which included various non-canonical G4 structures that are difficult to predict; conversely, G4s predicted by some search algorithms were not observed, highlighting the advantages of experimental G4 mapping. Subsequently, G4-seq reference maps have been made available for model organisms<sup>25</sup>. An analogous approach using reverse transcriptase stalling (rG4-seq) on poly(A)-enriched RNAs mapped RNA G4 structures in more than 3,000 human mRNAs<sup>26,27</sup> (FIG. 2b). These techniques provide important in vitro reference maps of G4-forming sites, although these data represent cell-population averages and do not account for the effect that proteins may have on G4 formation.

**Chemical mapping of G4s**

G4 chemical mapping exploits the different reactivity of nucleobases following the formation of G4 structures. Use of potassium permanganate-dependent single-strand nuclease (S1 nuclease) footprinting provided a snapshot of single-stranded DNA (ssDNA) regions of the genome in mouse B cell chromatin (FIG. 2c). The combination of ssDNA mapping with computational prediction of non-B-DNA sequence motifs identified putative G4 (and other non-B-DNA) structures in



**Fig. 2 | Approaches to detect and map DNA and RNA G-quadruplexes. a, b** | Mapping G-quadruplexes (G4s) by chain-extension stalling followed by high-throughput sequencing (G4-seq). G4s formed in the genome are mapped using G4-seq and G4s formed in the transcriptome are mapped using an analogous approach with reverse transcriptase stalling (rG4-seq). **a** | In G4-seq, a library of fragmented genomic DNA is sequenced twice, first in non-G4-forming conditions (Read 1) to provide a reference and then in G4-stabilizing conditions (for example, in the presence of  $K^+$  or the G4-stabilizing ligand pyridostatin (PDS); Read 2) to determine the positions of G4-dependent DNA polymerase stalling. **b** | In rG4-seq, poly(A)-enriched RNA is reverse transcribed in the presence of  $Li^+$  (a non-G4-forming condition) as a reference and with  $K^+$  or PDS to map RNA G4-dependent reverse transcriptase stalling. **c, d** | Chemical methods for mapping G4s. DNA G4s are mapped by potassium permanganate ( $KMnO_4$ )-single-strand nuclease (S1 nuclease) footprinting and RNA G4s are mapped by selective 2'-OH acylation analysed by primer extension (SHAPE). **c** | In  $KMnO_4$ -S1 nuclease footprinting,  $KMnO_4$  selectively oxidizes and traps single-stranded DNA (ssDNA), thereby allowing its digestion by S1 nuclease. Subsequent computational analyses infer the formation of DNA G4s based on the nuclease footprints. **d** | SHAPE utilizes differences in acylation kinetics of RNA 2'-hydroxyl groups treated with 2-methylnicotinic acid imidazole (NAI) and the ability of these groups to stall reverse transcription, to determine the formation of RNA G4s. **e** | Visualization of G4s (red foci) by immunofluorescence in the nucleus and cytoplasm using G4 antibodies (for example, the IgG antibody 1H6 or the single-chain variable fragment antibody (ScFv) BG4) together with fluorescently conjugated secondary or tertiary antibodies. **f** | PhenDC3 is an example of a fluorescence-labelled G4-targeting ligand. **g** | Mapping of DNA G4s using chromatin immunoprecipitation and next-generation sequencing (ChIP-seq) with G4-specific ScFv antibodies. BG4 precipitates DNA G4 structures from chromatin isolated using G4 ChIP-seq, whereas D1 is used in a ChIP-seq-like approach by expressing the antibody in cells. **h** | DNA G4 formation can be inferred indirectly by mapping the location of G4-binding proteins (G4BPs) using ChIP-seq.

gene regulatory regions<sup>28</sup>. Alternatively, the location of G4 structures can be deduced from their relative protection (compared with guanines in double-stranded DNA) from methylation by dimethyl sulfate (DMS) and subsequent cleavage by piperidine, which is provided by the Hoogsteen hydrogen interactions between the guanines in G-tetrads<sup>29</sup>. Another method of G4 chemical mapping is selective 2'-OH acylation analysed by primer extension (SHAPE; FIG. 2d), which exploits structure-dependent differences in acylation kinetics of RNA 2'-hydroxyl groups<sup>30</sup>. SHAPE performed in lithium versus potassium conditions was used to reveal RNA G4s *in vitro*<sup>31</sup>. Global analysis by DMS and SHAPE did not detect RNA G4s in eukaryotic cells<sup>26</sup>, prompting the hypothesis that RNA G4 structures are unfolded. However, caution is required when interpreting such data because of the limitations of chemical mapping, such as a map based on the averaging of dynamic structural states over time and across cell populations, along with the possibility of shifting the dynamic equilibrium of structural states during an experiment. These considerations have been discussed in more detail elsewhere<sup>32</sup>.

### Imaging G4s

G4s have been visualized by immunofluorescence in cells with G4 structure-specific antibodies (FIG. 2e). Use of the single-chain variable fragment antibody (scFv) Sty49 revealed G4 formation at the telomeres of a ciliate<sup>33</sup>. The scFv antibody BG4 revealed G4s in telomeric and non-telomeric DNA in fixed human cells<sup>34</sup>; this finding was corroborated by related observations using the G4-specific antibodies IgG 1H6 (REF.<sup>35</sup>) and scFv D1 (REF.<sup>36</sup>). BG4 was also employed to visualize RNA G4s in the cytoplasm of human cells<sup>37</sup>. An increased G4-antibody signal has been observed following the depletion of known G4-interacting proteins<sup>38–40</sup>, during S phase of the cell cycle and after treatment of live cells with G4-stabilizing ligands<sup>34,41</sup>. Although the specificity of each antibody may not be absolute<sup>42</sup>, and their sensitivity of detecting a single G4 structure as opposed to G4 clusters, is unproven in cells, there is an increasing level of cross-validation of G4 observations using these antibodies with observations of natural G4-interacting proteins.

Synthetic small molecules that recognize and stabilize G4s have also been used to probe cellular G4 structures. Derivatives of pyridostatin (PDS)<sup>43</sup> and PhenDC3 (REF.<sup>44</sup>) (FIG. 2f) incubated with live cells have subsequently been conjugated to a fluorescence probe using 'click chemistry' after cell fixation to visualize G4s. G4 ligands with intrinsic fluorescence, such as DAOTA-M2, support the existence of G4s in cells<sup>45</sup>. Real-time, live cell imaging provided by such ligands has advantages over antibodies, which require cell fixation and permeabilization, although in live cells the sensitivity of fluorescence to G4-independent changes in pH, redox status, intracellular polarity and viscosity require careful consideration to validate G4-specific binding.

### Genomic and transcriptomic techniques

DNA G4s have been detected and mapped in the chromatin of human cells using chromatin immunoprecipitation followed by high-throughput sequencing (G4

ChIP-seq) using BG4 (REFS<sup>46,47</sup>) (FIG. 2g). Strong overlap between these data and the human G4-seq reference map (of all genomic sequences that can form a G4) can provide cross-validation<sup>24</sup>, as does a G4 map generated from the expression of the G4-specific D1 antibody in cells<sup>36</sup> (FIG. 2g). Alternatively, G4s can be inferred using antibodies against known G4-binding proteins. For example, the helicases  $\alpha$ -thalassaemia mental retardation X-linked protein (ATRAX)<sup>48</sup>, xeroderma pigmentosum group B (XPB) and XPD<sup>49</sup> bind folded G4 oligonucleotides *in vitro*, and have been mapped to G4 motifs in human chromatin using ChIP-seq (FIG. 2h). In yeast, the G4-associated ATP-dependent DNA helicase Pif1 (REF.<sup>50</sup>), ribosome biogenesis protein SLX9 (REF.<sup>51</sup>) and Rap1-interacting factor 1 (Rif1)<sup>52</sup> have been mapped to G4 motifs. Limitations of ChIP-seq include the averaging of results obtained from cell populations, the inability to resolve temporal dynamics and the dependence on accessibility of the target to the antibody. There are also potential biases introduced by antibody specificity, sample fixation and fragmentation that must be eliminated by carefully designed control experiments<sup>53</sup>. Single-cell genomics approaches, super-resolution microscopy and live cell imaging by single-molecule light sheet microscopy may, in due course, provide additional insights into the nature and dynamics of DNA G4s in cells.

Recent studies have used RNA immunoprecipitation techniques, such as individual-nucleotide resolution ultraviolet crosslinking and immunoprecipitation (iCLIP) and photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (parCLIP), to identify transcriptome-wide binding sites of known RNA G4 protein interactors<sup>54–56</sup>. These techniques provide a valid strategy for identifying RNA G4s that exist in a protein-bound state and the biological processes they influence. However, they do not provide explicit proof of the existence of an RNA G4-folded structure at these sites.

G4 occurrence in living cells has now been demonstrated using chemical mapping, antibody-based or small molecule-based imaging and sequencing techniques. The endogenous landscape of G4s, revealed by these methods, is only a small fraction of the total number of possible G4s in the genome. Although this may be a fair reflection of biological reality, it remains possible that existing technologies do not detect the full repertoire of G4s in cells. In particular, transient and/or dynamic G4s states may not be accurately detected with current approaches (TABLE 1 summarizes the advantages and limitations of current technologies). The development of new tools and methodologies, alongside technical advances in live cell imaging and genome editing, will provide further insights in the future.

### Control of G4 formation and unwinding

Sequences that form G4s are prevalent in genomes, and it is becoming clear that not all sequences with G4 potential form structures in cells and that different cell types or cell states have distinct patterns of G4 formation. Understanding how G4 formation is regulated in a cellular context is therefore a question of fundamental importance. Biophysical approaches have been used to

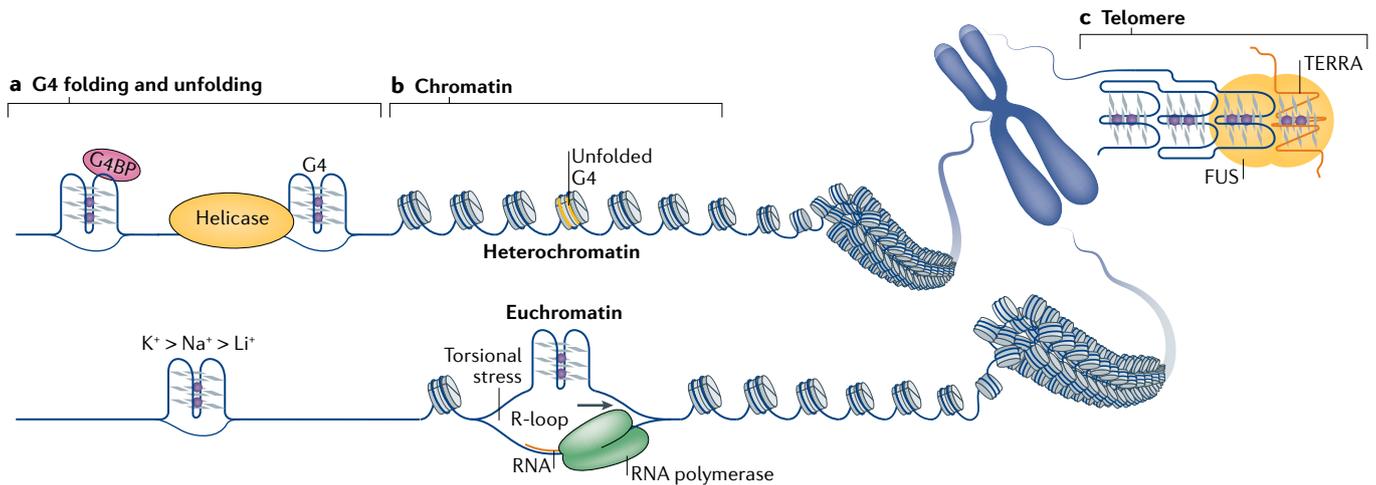
Table 1 | Advantages and limitations of techniques to map and detect G-quadruplexes (G4s)

| Methodology                         | Technique                                                                                 | Uses and advantages                                                                                                                                                              | Limitations                                                                                                                                                                                                                                                                                                                                                                                                             |
|-------------------------------------|-------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Mapping by chain-extension stalling | G4-seq                                                                                    | Identifies in vitro nucleic acid sequences with the potential to form G4s in the genome (G4-seq) or transcriptome (rG4-seq)                                                      | Performed on extracted DNA or RNA. Thus, the influence of the cellular environment, for example proteins or chromatin structure, on the G4 landscape is not considered                                                                                                                                                                                                                                                  |
|                                     | RNA G4-seq (rG4-seq)                                                                      |                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                         |
| Chemical mapping                    | Potassium permanganate–S1 nuclease footprinting                                           | Maps genome-wide multiple types of non-B DNA structures in the chromatin context                                                                                                 | Relies on motif-annotation algorithms to map the non-B-DNA structures (including G4 formation)<br>Cannot accurately discern individual non-B DNA structures at sites containing large clusters of non-B DNA<br>Readout is an averaging of structural states<br>Can shift the dynamic equilibrium of structural states and hence may not reflect true intracellular structures. Readout is not specific to G4 structures |
|                                     | Selective 2'-OH acylation analysed by primer extension (SHAPE)                            | Provides quantitative, single-nucleotide-resolution RNA structural information                                                                                                   | Structural information is lost at both the 5' and 3' ends of an RNA because the technology depends on primer extension<br>Readout is not specific to G4 structures                                                                                                                                                                                                                                                      |
|                                     | In vivo dimethyl-sulphate (DMS) footprinting                                              | Determines nucleic acid (DNA and RNA) secondary and tertiary structures at single-nucleotide resolution<br>DMS easily and rapidly penetrates cells and all cellular compartments | Readout is an averaging of structural states<br>High cellular toxicity<br>Can shift the dynamic equilibrium of structural states and, hence, does not reflect true cellular structures<br>DMS reactivity depends on solvent accessibility and local electrostatic environment                                                                                                                                           |
| Antibody-based mapping              | G4 chromatin immunoprecipitation sequencing (G4 ChIP-seq)                                 | Genome-wide mapping of DNA G4s in the chromatin context                                                                                                                          | Possible biases introduced by sample fixation and fragmentation<br>Relies on antibody specificity, target accessibility and cell-population averaging                                                                                                                                                                                                                                                                   |
|                                     | ChIP-seq of G4-binding proteins                                                           | Genome-wide mapping of G4 DNA binding proteins in the chromatin context                                                                                                          | Cannot determine on which DNA strand G4s are located<br>Antibodies against G4-binding proteins provide indirect evidence of DNA G4s, which relies on the specificity of the protein for G4s                                                                                                                                                                                                                             |
|                                     | Individual-nucleotide resolution ultraviolet crosslinking and immunoprecipitation (iCLIP) | Identifies all RNA sequences bound to the RNA binding protein (RBP) of interest                                                                                                  | Relies on the specificity of the RBP to bind RNA G4<br>Cannot account for protein binding to unfolded G4 sequence motifs<br>Relies on cell-population averaging                                                                                                                                                                                                                                                         |
| Imaging                             | Immunofluorescence                                                                        | Single-cell resolution of G4 abundance<br>Possible to detect DNA and RNA G4s simultaneously                                                                                      | Requires cellular fixation and permeabilization<br>Relies on the specificity of the antibody<br>Does not provide sequence context<br>Undetermined resolution: do detected G4 foci represent one or several G4s?                                                                                                                                                                                                         |
|                                     | Fluorescent G4-stabilizing ligands                                                        | Allow the study of dynamic formation of G4s in fixed and live cells                                                                                                              | Fluorescence is sensitive to cellular changes in pH, polarity (hydrophilic versus hydrophobic compartments) and viscosity, making discrimination of G4-specific from non-specific binding events difficult<br>The dynamic equilibrium of G4 formation may be shifted by the experiment and thus will not reflect the true cellular state<br>Lack of sequence context<br>Relies on ligand specificity and half-life      |

G4-seq, genome-wide DNA polymerase-stop assay followed by high-throughput sequencing; rG4-seq, transcriptome-wide reverse transcriptase stalling assay followed by high-throughput sequencing; S1 nuclease, a single-strand nuclease.

evaluate properties that influence DNA and RNA G4 formation in oligonucleotides in near-physiological conditions<sup>57</sup>. G4 stability is affected by numerous factors, including the number of G-tetrads, loop length and topology<sup>58,59</sup> and the sequence composition, both within the G4 motif and flanking regions<sup>19</sup>. RNA G4s are generally more thermally stable than their DNA counterparts<sup>57</sup>. G4s are stabilized by centrally located,

monovalent cations in the order  $K^+ > Na^+ > Li^+$  (REF.<sup>60</sup>) (FIG. 1), which may be physiologically relevant because  $K^+$  is the most abundant metal ion in mammalian cells<sup>61</sup>. Furthermore, G4 formation can be favoured by the induction of negative torsional stress behind RNA and DNA polymerases<sup>62</sup> and by molecular crowding<sup>63</sup>, which are both relevant in the cellular context of genomic DNA.



**Fig. 3 | Regulation of G-quadruplex structure formation.** **a** | Physical factors, such as the presence of stabilizing cations (bottom), length and sequence composition (not shown) of the loops and flanking sequences, determine the thermodynamic stability of G-quadruplexes (G4s). In cells, specialized proteins that unwind G4s (for example, helicases) or bind and stabilize G4s (for example, some G4-binding proteins (G4BPs)) can shape the G4 landscape. **b** | The genomic structural context, such as the negative superhelicity behind the RNA polymerase complex, as well as the crosstalk with other structural phenomena, like RNA–DNA hybrids known as R-loops that arise from hybridization of the nascent RNA with the template DNA, can contribute to G4 formation (bottom). Chromatin structure seems to have a strong influence on G4 formation, as the majority of endogenous G4s have been mapped to nucleosome-depleted, open chromatin regions. **c** | Telomere heterochromatin homeostasis can be influenced by the simultaneous binding of proteins, such as fused in sarcoma (FUS), to G4s formed in telomere DNA and in the long non-coding RNA TERRA (telomeric repeat RNA).

#### Helicases and other G4-binding proteins

G4s can impede nucleic acid functions (for example, DNA replication, transcription or translation) and proteins exist that can resolve them (FIG. 3a). Many such proteins belong to canonical helicase families, such as the RecQ-like and DEAD box or DEAH box helicase families. The RecQ-like helicases Bloom syndrome protein (BLM) and Werner syndrome ATP-dependent helicase (WRN) were the first recognized G4-resolving mammalian helicases<sup>64,65</sup>. DEAH-box helicase DHX36 is able to resolve both RNA and DNA G4s<sup>66</sup>. In vitro, these three helicases unwind G4s formed within oligonucleotides by binding the 3′ tail-containing DNA substrate and performing a repetitive 3′–5′, ATP-independent unfolding of the structure<sup>67–69</sup>. In the case of BLM and WRN, this activity is a result of cooperative binding of their helicase–RNaseD domain to the 3′ ssDNA and their RecQ domain to the G4<sup>67</sup>. For DHX36, G4 binding induces rearrangements in the helicase core, which pulls the single-stranded region, thereby tugging the G4 one nucleotide out of alignment<sup>68</sup>. This is sufficient to destabilize a DNA G4 in the presence of its complementary strand<sup>69</sup>. DHX36 is also able to fully unwind an RNA G4, but then refolds it in the absence of a complementary strand, giving rise to dynamic G4 unfolding and refolding<sup>70</sup>. Pif1 (in budding yeast) and Fanconi anaemia group J (FANCI) unwind G4s 5′–3′ in an ATP-dependent manner, requiring (in vitro) a 5′ tail-containing DNA<sup>71,72</sup>.

Other non-helicase proteins, such as cellular nucleic acid-binding protein (CNBP)<sup>73</sup> and G-rich sequence factor 1 (REF.<sup>74</sup>), have been reported to sequester unfolded form of G4s. The ssDNA-binding proteins protection of telomeres 1 (POT1) and replication protein A (RPA) use a Brownian ratchet-like mechanism and unfold G4s

in a multistep process<sup>75</sup>. However, additional mechanisms such as simple trapping of the unfolded G4 have also been proposed for POT1 (REF.<sup>76</sup>), suggesting that its mechanism of function remains somewhat elusive. Conversely, proteins such as nucleolin<sup>77</sup> and LARK (also known as RNA-binding protein 4 (RBM4))<sup>78</sup> were shown to stabilize G4 structures. Several studies have applied affinity enrichment, using G4 baits, to identify G4 interacting proteins, most of which require further studies to elucidate the nature and biological relevance of the interaction<sup>54,79</sup>. For many proteins identified to bind G4 oligonucleotides, evidence for specific recognition of the G4 structure in vivo is lacking.

#### Local features favouring DNA G4s

G4 formation in genomic DNA competes with Watson–Crick base-pairing; however, this base-pairing is necessarily disrupted during replication, transcription and DNA damage repair, thereby favouring G4 formation. During transcription, negative torsional stress induced behind the RNA polymerase complex can be relayed upstream (in the opposite direction of transcription) and promote melting of duplex DNA, which contributes to G4 formation, for example as proposed for the far upstream element of the human *MYC* promoter<sup>80</sup> (FIG. 3b). The formation of RNA–DNA hybrids known as R-loops from the hybridization of the nascent RNA with the template DNA may contribute to G4 formation on the displaced DNA strand, as observed by electron microscopy<sup>81</sup>. G4s and R-loops are favoured by similar DNA features, such as GC-richness and negative torsional tension, and genome-wide profiling of R-loops in human embryonic kidney cells revealed considerable overlap with G4-forming sequences identified by

**Polytene chromosomes**

Giant chromosomes found in particular tissues of various eukaryotes, which are formed following several rounds of DNA replication without cell division.

**Fragile telomeres**

Aberrant or discontinuous appearance of telomere chromatin in metaphase chromosomes, identified by fluorescence in situ hybridization and indicative of telomere replication defects.

G4-seq<sup>82</sup>. C-rich DNA can form an intercalated motif (i-motif) secondary structure by stacking intercalated and hemi-protonated cytosine base pairs (C<sup>+</sup>:C)<sup>83</sup>. Biophysical studies have shown that i-motifs are favoured by acidic pH, although recent antibody-based experiments have shown i-motif formation in nuclei of fixed human cells<sup>84</sup>. Although i-motifs can in principle occur on the C-rich strand opposite G4, cell cycle analysis shows that i-motif formation is maximal in late G1, whereas G4 formation peaks in S phase, indicating that the situation may be more complex<sup>85</sup>.

Studies to detect and map DNA G4s in a chromatin context have shown that G4s occur primarily at regulatory, nucleosome-depleted regions and promoters of actively transcribed genes in human cancer cells using G4 ChIP-seq<sup>46,47</sup> (FIG. 3b) and in mouse B cells using permanganate footprinting<sup>28</sup>. Furthermore, in human cells, G4s co-localized with RNA polymerase II (Pol II) and with trimethylated histone H3 Lys4 (H3K4me3), which is a histone modification associated with active genes but not with the heterochromatin modification H3K9me3 (REF.<sup>46</sup>). Overall, G4 structures were detected in relatively accessible chromatin<sup>46,47</sup>. By contrast, in *Drosophila melanogaster* polytene chromosomes, G4s were found in the heterochromatin<sup>86</sup>, suggesting the existence of species-specific differences in G4 formation.

**Oxidative base damage and G4 formation.** G4s are sensitive to redox chemistry and early work showed that G4s in complex with porphyrins, such as haem, have peroxidase and peroxygenase enzyme-like activity<sup>87</sup>. Free haem is potentially toxic as it can catalyse the formation of reactive oxygen species, leading to oxidative stress. It has thus been hypothesized that G4s act as a sink for free haem to prevent DNA damage<sup>88</sup>. Reactive oxygen species can induce DNA damage in the form of 8-oxoguanine (8-oxoG), particularly in G-rich regions. For example, at telomeres, 8-oxoG disrupts G4 formation, stimulates telomerase function and promotes telomere instability<sup>89</sup>. Considerable transcriptional and DNA damage responses are also observed following oxidative damage at promoter G4s, such as in the vascular endothelial growth factor (*VEGF*; also known as *VEGFA*) gene<sup>90</sup>.

**Cellular functions of G4s**

A major and largely unanswered question is what G4s do. G4s are found in so many different cellular contexts that they can be considered either physical obstacles that must be overcome to enable some nucleic-acid related process or useful for normal cellular functions.

**G4 structures at telomeres**

DNA and RNA G4s have roles in telomere biology. Telomeres are nucleoprotein structures located at chromosome ends, which maintain genome integrity by suppressing aberrant DNA repair of the DNA ends through binding of telomere-specific protein complexes and formation of higher-order DNA secondary structures<sup>91</sup>. In the somatic nucleus of the ciliate *Stylonychia lemnae*, protection is provided by intermolecular telomere DNA G4s, which are stabilized by telomere-binding protein- $\alpha$

(TEBP $\alpha$ ) and TEBP $\beta$ <sup>92</sup>. In yeast and vertebrates, protection is provided by the 'lasso-like' telomere loop, in which the telomere ssDNA overhang invades the upstream double-stranded telomere DNA<sup>93</sup>. G4s may also cap chromosome ends, as G4s have been detected at telomeres<sup>34–36</sup> and several telomeric proteins (for example, RIF1 (REF.<sup>52</sup>) and telomeric repeat-binding factor 2 (TRF2)<sup>94</sup>) bind G4s in vitro. Indeed, G4s can act as rudimentary protective structures when the normal telomere capping structure is compromised<sup>95</sup>. Chromatin homeostasis at telomeres and at sub-telomeric regions is also dependent on the long non-coding RNA TERRA (telomeric repeat RNA) forming a G4 structure, which was proposed to be a protein docking scaffold<sup>96</sup> (FIG. 3c). The human proteins TRF2, Ewing sarcoma breakpoint region 1 protein (EWS) and fused in sarcoma (FUS) can co-bind the TERRA G4 and telomere DNA G4 in vitro<sup>94,96,97</sup>. By co-binding TERRA G4 and telomere G4, FUS can recruit histone methyltransferases that are important for telomere and sub-telomere heterochromatin maintenance<sup>96</sup>.

Persistent formation of G4s at telomeres during DNA replication is problematic. Depletion of many proteins known to interact with telomeric G4 oligonucleotides (for example, the CTC1–STN1–TEN1 (CST) complex<sup>39</sup> and helicases like regulator of telomere elongation helicase 1 (RTEL1)<sup>98</sup>) results in telomere shortening, altered telomere replication rate and/or formation of fragile telomeres, which arise from stalling of replication forks at telomeres during lagging strand synthesis<sup>99</sup>. Furthermore, these phenotypes are exacerbated by the presence of G4-stabilizing ligands; for example, sudden telomere loss occurs only in cells subjected to the combination of CST depletion and PDS treatment<sup>39</sup>.

G4 formation can control access to telomeres of telomerase, the non-coding RNA–reverse transcriptase complex that extends 3' ends of chromosomes in cancer cells, stem cells and cells of the germline to prevent telomere shortening and genome instability. Formation of anti-parallel intramolecular telomere DNA G4s (FIG. 1d), prevents telomere extension by limiting access to the 3' end of the telomere to telomerase<sup>100</sup>, whereas parallel intermolecular telomere DNA G4s can be extended due to partial G4 resolution by telomerase in vitro<sup>101</sup>. Supporting these findings, in *Saccharomyces cerevisiae*, parallel telomeric G4 stabilization by the telomere elongation protein Est1 is essential for telomerase recruitment<sup>102</sup>. Furthermore, in S phase, when G4s can co-localize with human telomerase, intermolecular G4s between sister chromatids have been hypothesized to form<sup>101</sup>. However, the complex POT1–tripeptidyl-peptidase 1 (TPP1), which is responsible for the processivity and recruitment to telomeres of human telomerase, is known to destabilize G4 structures<sup>76,103</sup>. The recent observation in vitro that G4 folding within the active pocket of human telomerase supports POT1–TPP1-dependent telomerase processivity<sup>104</sup> suggests that G4 formation is important during human telomere DNA synthesis. Telomerase is also influenced by the unfolding of a G4 at the 5' end of the RNA component of telomerase by DHX36 (REF.<sup>105</sup>).

Although the natural function of telomere G4s with regards to telomerase is still unclear in vivo,

ligand-stabilized telomere G4s inhibit telomerase-mediated telomere extension<sup>106</sup>. The additional ability of G4 ligands to displace components of the telomere protection complex shelterin (for example, TRF2 and POT1) results in telomere DNA damage and cell death, and has led to the development of a plethora of G4 ligands as potential chemotherapeutic agents<sup>9,107</sup>. Furthermore, although G4-induced replication stress in certain cancer cells has been proposed to promote alternative lengthening of telomeres, which is a mechanism activated in 15% of cancers<sup>108</sup>, treatment with G4 ligands is still effective in killing these cells<sup>109</sup>, thereby supporting the consideration of G4 ligands as pan-anticancer compounds.

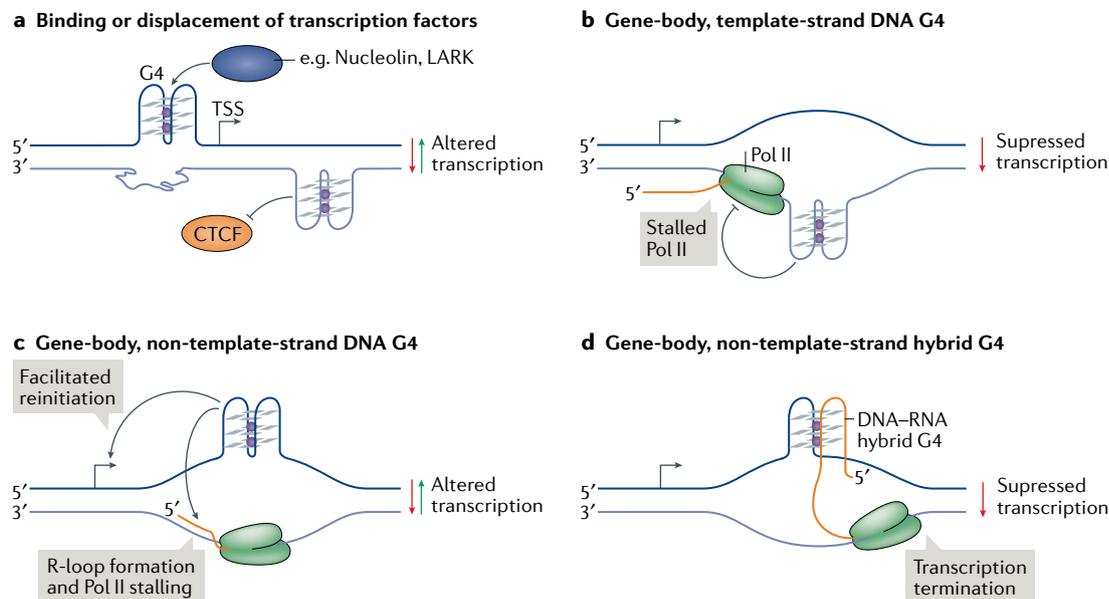
**Transcription**

A K<sup>+</sup>-dependent G4 structure was initially detected in the promoter of the chicken β-globin-encoding gene<sup>21</sup>, and subsequently other G4-forming sequences were noted in the promoters of human genes, most notably in *MYC* and several other oncogenes<sup>110</sup>. Computational predictions indicated that G4 motifs are prevalent and enriched in human gene promoters compared with the rest of the genome<sup>13</sup>, suggesting that G4s are involved in transcription regulation. Notably, sequencing-derived genomic G4 maps of 12 organisms confirmed increased G4-forming potential at gene promoters and 5' untranslated regions (5' UTRs) in human, mouse and *Trypanosoma brucei*, which indicates the existence of functional similarities in their G4 biology<sup>24,25</sup>. By contrast, no enrichment or even depletion was found in lower eukaryotes and in bacteria. Recent studies have mapped G4 structures in chromatin to regulatory regions upstream of the transcription start sites of

actively transcribing genes in human cells<sup>28,46</sup>, further supporting the link between G4s and transcription.

Many publications report that transcription can be regulated by small molecules that target G4s. Treatment of cells with the G4-stabilizing ligand TMPyP4 resulted in reduced *MYC* expression<sup>111</sup>, with comparable observations subsequently reported for other gene promoters, such as *KRAS*<sup>112</sup> and *KIT*<sup>113</sup>. Furthermore, transcriptome-wide changes have been reported at genes with promoter G4s<sup>114</sup>. Although such correlations are consistent with the existence of a G4–transcription link, more explicit evidence of small-molecule binding to folded promoter G4s in cellular DNA, along with an improved consideration of the potential consequences of indirect, network effects on transcription, would help to better characterize the link.

Computational analysis of the binding motifs of several transcription factors showed that they are strongly enriched in certain promoter G4 motifs<sup>115</sup> and in endogenous G4s as observed by G4 ChIP-seq<sup>116</sup>. This enrichment suggests that positive or negative interactions exist between transcription factors and G4 structures at gene promoters. An early study reported the binding in vitro of the G4 motif from the *MYC* promoter by recombinant nucleolin<sup>77</sup> (FIG. 4a). By contrast, Förster resonance energy transfer experiments indicated the unfolding of a *MYC* G4 oligonucleotide upon binding of NM23-H2 (also known as NDK-B)<sup>117</sup>, prompting the hypothesis that nucleolin and NM23-H2 are involved in G4 stabilization and unwinding, respectively, to regulate *MYC* transcription. Since then, in vitro interactions with G4 oligonucleotides have been shown for several other transcription factors, including CNBP<sup>118</sup>, SP1 (REF.<sup>119</sup>)



**Fig. 4 | Models of G-quadruplex involvement in transcription. a** | DNA G-quadruplexes (G4s) upstream of the transcription start site (TSS) or in the gene body could bind or displace transcription factors, resulting in altered transcription. **b** | During transcription elongation, the separation of DNA strands in the transcription bubble may result in the formation of G4s in gene bodies. G4 formation on the template strand can block the progression of RNA polymerase II (Pol II). **c** | Gene-body G4s on the non-template strand may facilitate transcription re-initiation. Conversely, such G4s may favour the stable association of nascent RNA (orange) with the template DNA, resulting in the formation of RNA–DNA hybrids known as R-loops and in Pol II stalling. **d** | Formation of DNA–RNA hybrid G4s between the non-template DNA and the nascent RNA can lead to premature transcription termination.

and LARK<sup>78</sup>. Additional experiments confirming G4 structure formation as part of protein–DNA complexes, ideally using structural biology, and explicit evidence for binding at endogenous G4s would help strengthen the link between protein binding and G4 status. Conversely, G4 formation in the first exon of the human telomerase reverse transcriptase (*hTERT*) gene has been suggested to disrupt binding of the gene repressor CCCTC-binding factor, resulting in elevated plasmid-encoded *hTERT* transcription<sup>120</sup> (FIG. 4a).

Several studies have investigated the effects of G4 motifs in gene bodies on transcription elongation. A substantial inhibitory effect was observed when a G4 motif was present on the template strand, both in human embryonic kidney cells<sup>121</sup> and in *Escherichia coli*<sup>122</sup>, which is consistent with impairment of Pol II progression by G4 structures in the template strand (FIG. 4b). By contrast, G-rich sequences inserted in the non-template strand (but not in the template strand) impaired in vitro transcription by T7 polymerase<sup>123</sup>. In this case, transcription blockage was ascribed to R-loop formation independent of G4 formation, as no major transcriptional differences were observed between G4-stabilizing and G4-destabilizing buffer conditions. Computational analysis of human genes has observed a correlation between increased promoter-proximal Pol II pausing and the presence of downstream G4 motifs on the template and non-template strands<sup>124</sup>. In addition, human genes with a greater number of G4s on the non-template strand up to 500 base pairs downstream of the transcription start site are associated with higher than average steady-state transcription levels and Pol II occupancy, suggesting that G4s on the non-template strand could maintain the DNA in an open state and, thus, aid transcription reinitiation<sup>125</sup> (FIGS 3b,4c).

Transcription of the mitochondrial gene *CSB II* was investigated in vitro using 7-deaza-dGTP or 7-deaza-GTP nucleotides, which cannot form Hoogsteen base-pairing and thus cannot stabilize G4s. The nascent RNA and non-template DNA strand were shown to co-transcriptionally form a stable DNA–RNA hybrid G4, which was suggested to promote transcription termination<sup>126</sup> (FIG. 4d). Similarly, transcription suppression was observed when a hybrid-G4 forming sequence was inserted into reporter plasmids<sup>127</sup>.

Models linking G4s and transcription have been largely based on computationally predicted G4 sequences, supporting correlations and the manipulation of isolated G4 structures in plasmid constructs. It is also evident that the specific G4 positions (for example, in regulatory regions or gene bodies and in template strand or non-template strand) may contribute to different regulatory mechanisms. Furthermore, the local chromatin context appears to have a substantial effect on G4 formation and function<sup>46</sup>. Additional work is needed to elucidate the molecular mechanistic details of how G4s influence transcription in a chromatin context.

#### The effects of G4s on genome stability

G4s can cause replication stress by obstructing the progression of DNA replication forks and causing replication-fork collapse<sup>128,129</sup>, which generates DNA

double-strand breaks that can lead to genome instability. Computational analyses of large cancer datasets associated G4s with breakpoints that accompany somatic copy-number alterations<sup>130</sup>. Another large cancer association study found that G4 motifs, particularly thermodynamically more stable G4s, were enriched at sites of somatic mutations, implying that G4 structures increase the probability of recurrent mutations and may be important determinants of mutagenesis<sup>131</sup>. The link between G4s and genome instability has been strengthened by sequencing of G4s in the human genome, which revealed a notable association of G4s with gene amplifications commonly observed in cancers<sup>24,46</sup>.

Studies in model organisms provided substantial support in vivo for G4s being a direct cause of genome instability. *Caenorhabditis elegans* lacking *dog-1* (also known as helicase ATP-binding domain-containing protein), which is the ortholog of the helicase FANCI, accumulate deletions at G-rich regions, including in predicted G4 motifs<sup>132</sup>. Experiments using plasmids as replication templates in *Xenopus laevis* egg extracts showed that the absence of FANCI or the presence of G4-stabilizing ligands cause replication stalling at G4 structures<sup>133</sup>. In *C. elegans*, genetic analyses have also demonstrated that site-specific genome deletions can originate from a single predicted G4 sequence motif<sup>134</sup>. Genetic analyses in *S. cerevisiae* have also shown that Pif1, a potent G4-unwinding helicase, suppresses DNA damage and gross chromosomal rearrangements mediated by G4s<sup>135</sup>. Human minisatellite tandem repeats comprising G4 motifs also show increased instability when introduced into *S. cerevisiae* lacking Pif1 or in the presence of G4-stabilizing ligands<sup>136,137</sup>; thermodynamically stable G4s with short loops preferentially caused rearrangements<sup>138</sup>.

**Helicases prevent G4-induced genome instability in humans.** G4 helicases protect the genome by unfolding G4s that can cause DNA breakage and subsequent aberrant recombination; failure to resolve G4s owing to loss of helicase activity may induce genome instability. Sister chromatid exchanges are common in cells of individuals with Bloom syndrome and are enriched at predicted G4 sites, particularly in transcribed genes<sup>139</sup>. In glioma cells, ATRX loss promotes G4 formation, somatic copy-number alterations and increased occupancy of BLM at DNA damage sites<sup>38</sup>. Chromosomal regions known as common fragile sites are predisposed to breakage and undergoing rearrangements during replication stress. ATRX localizes to common fragile sites during replication stress and ATRX loss is associated with increased numbers of chromosomal breaks at common fragile sites<sup>140</sup>. FANCI together with the ssDNA-binding protein RPA enable S-phase progression by facilitating G4 unwinding<sup>72</sup>. BLM and WRN also form complexes with RPA, which are mediated by the BRCA1-interacting E3 ubiquitin-protein ligase HERC2 (REF. 40). Similar to BLM and WRN loss, HERC2 depletion or inhibition of its ubiquitylation activity increases G4 formation. In a functional genomic screen for G4-interacting factors, HERC2 loss was found to promote cell death in cells treated with G4-stabilizing ligands<sup>141</sup>. BLM and WRN may operate independently or on different G4s in the

#### Common fragile sites

Specific chromosomal regions that are intrinsically hard to replicate and preferentially form chromatin gaps or breaks during metaphase following replication stress.

genome, whereas HERC2 seems to be epistatic to BLM and WRN and a master regulator of G4 suppression<sup>40</sup>.

**G4 stabilization induces DNA damage.** Small molecules, such as PDS, can stabilize G4s<sup>142</sup> and cause replication-dependent and transcription-dependent DNA double-strand breaks (detected by  $\gamma$ -H2AX ChIP-seq), which map to G4-rich regions at loci that include several oncogenes<sup>43</sup>. Cells compromised in their ability to process G4s are particularly sensitive to G4-stabilizing ligands. For example, loss of FANCI, HERC2 or ATRX sensitizes cells to different G4 ligands, including to telomestatin, PDS and CX-3543 (REFS<sup>38,40,72</sup>). In mice, ATRX-deficient glioma xenografts are growth impaired by CX-3543 and the host mice show increased survival<sup>38</sup>, highlighting the therapeutic potential of G4 ligands.

**G4s and R-loops.** R-loops and G4s can form on opposite DNA strands (FIGS 3b,4c), and then threaten genome stability by blocking DNA and RNA polymerases and causing transcription–replication conflicts. Immunoprecipitation of such DNA–RNA hybrids with an antibody or using an inactive version of the R-loop processing enzyme RNase H1 identified R-loops in GC-rich promoter regions<sup>143,144</sup>, including regions enriched in G4 motifs<sup>82</sup>. It is striking that an immediate response to G4 stabilization by PDS is an increase in R-loops, particularly opposite G4 sites, which results in DNA damage followed by the formation of micronuclei<sup>145</sup>. Although the underlying mechanism of this process is unclear, overexpression of RNase H1 counteracts these effects of G4 ligands, suggesting that R-loops are required for the induction of DNA damage by G4 stabilization<sup>145</sup>. G4 and R-loop induction by PDS were shown in HeLa and U2OS cancer cell lines, but are absent in immortalized fibroblasts, suggesting that aberrant G4-related pathways are present in cancer cells. Notably, ATRX loss also results in increased R-loop formation at telomeres<sup>146</sup>, consistent with coupling of G4 regulation with R-loop stability.

### Translation

G4-mediated translation inhibition was first reported for an RNA G4 from the coding region of the mRNA of fragile X mental retardation protein (FMRP; also known as synaptic functional regulator FMR1); when inserted into a luciferase reporter, this RNA G4 caused a 1.5-fold reduction in translation *in vitro* in reticulocyte lysates<sup>147</sup>. A similar *in vitro* reporter assay showed that a G4 from the 5' UTR of the NRAS mRNA caused a fourfold reduction in translation<sup>148</sup>. Subsequently, G4s from numerous 5' UTRs, including the mRNAs of BCL2 (REF. 149) and ADAM10 (REF. 150), were shown to inhibit translation in cell-free or cell-based reporter assays. The G4-forming triplet repeat expansion (CGG)<sub>60–90</sub> from the 5' UTR of the *FMR1* gene also inhibits translation of a reporter construct and of exogenously expressed FMR1 mRNA in cells, but, despite folding into G4 structures, the shorter (CGG)<sub>30</sub> repeats elevate translation<sup>151</sup>. RNA G4 density, thermodynamic stability and position relative to the 5' cap have all been shown to differentially influence translation<sup>151,152</sup>.

G4 motifs are often located near the beginning of 5' UTRs, suggesting they have a role in translation

initiation<sup>153</sup>. The helicase eukaryotic initiation factor 4A (eIF4A) unwinds structured 5' UTRs to facilitate the recruitment of the 43S pre-initiation complex and the subsequent scanning for the start codon<sup>154</sup>. Depletion of eIF4A or its inhibition by silvestrol reduced the translation efficiency of mRNAs with longer 5' UTRs enriched in two-tetrad G4s, indicating that RNA G4s directly influence recruitment of, or scanning by, the ribosome<sup>155,156</sup>. Unresolved G4s in 5' UTRs can promote the formation of 80S ribosomes on alternative, upstream start codons, thus inhibiting the translation of the main open reading frame<sup>55</sup> (FIG. 5a). RNA G4s in the mRNAs of FGF2 (REF. 157),  $\alpha$ -synuclein<sup>158</sup> and VEGF<sup>159</sup> stimulate internal ribosome entry site (IRES)-mediated translation, potentially by recruiting the 40S ribosome<sup>160</sup>. However, RNA G4s in the IRES of VEGF and in  $\alpha$ -synuclein were subsequently found to be functionally dispensable, and the role of RNA G4s in IRES-dependent translation is far from being clear<sup>158,161,162</sup>.

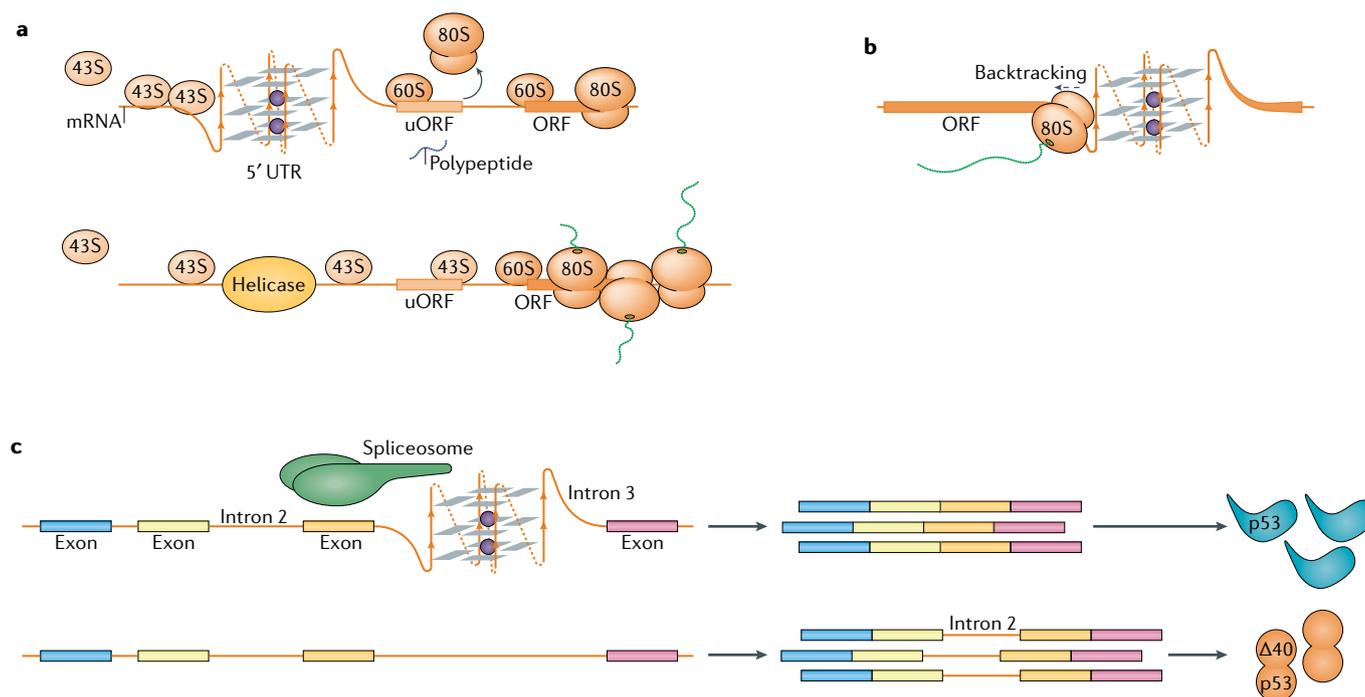
G4s occur at a lower abundance in mRNA coding sequences when compared with 5' UTRs<sup>153</sup>. In the coding sequence, ribosomes stall 6–7 nucleotides before a G4, which is the distance of the tRNA acceptor site from the RNA entry site<sup>163</sup>. Thus, despite the helicase activity associated with the 80S ribosome, RNA G4s are problematic for translation elongation and, therefore, evolutionarily selected against through the use of synonymous codons<sup>164</sup> (FIG. 5b).

As indicated above, interactions of RNA G4s with RNA-binding proteins (RBPs) can influence translation. A prime example is FMRP, which regulates translation of hundreds of mRNAs and is reviewed in detail elsewhere<sup>165</sup>. Other examples include CNBP, which binds RNA G4s in 4,178 different mRNAs and elevates translation by preventing the formation of RNA G4s<sup>73</sup>, and DHX36, depletion of which reduces mRNA translation<sup>56</sup>.

Our current knowledge of the effects of RNA G4s on translation relies substantially on reporter assays, which are a helpful tool, but care must be taken when interpreting results obtained from such non-native systems. For example, an isolated G4 derived from the 5' UTR of the TGF $\beta$ 2 mRNA reduces translation when embedded in a reporter mRNA, but conversely enhances translation in the context of the entire UTR<sup>166</sup>. In another example, a G4 from BCL2 mRNA was shown to suppress translation *in vitro*, but its genomic disruption by CRISPR-Cas9 failed to increase translation of the native mRNA<sup>167</sup>. Interestingly, translation output is unaltered when G4s from the mRNAs of VEGF and TGF $\beta$ 2 associated with translation stimulation are replaced with G4s associated with translation repression from the mRNAs of NRAS and MT3-MMP, and vice versa<sup>168</sup>. Therefore, the context in which RNA G4s form, their dynamic relationship with surrounding alternative structures<sup>169,170</sup> and their interactions with RBPs should all be considered when evaluating translational output.

### Other roles of G4s in RNA biology

RNA G4s can influence the subcellular localization of mRNAs in neurons<sup>171</sup>. Interaction of RNA G4s with RBPs, such as FMRP<sup>172</sup>, regulates their localization and local translation in dendrites. Recent studies have



**Fig. 5 | G-quadruplexes in RNA biology.** **a** | Formation of RNA G-quadruplexes (G4s) impedes scanning of the 5' untranslated region (UTR) by 43S ribosomes and leads to translation initiation at an upstream open reading frame (uORF) at the expense of translation of the main ORF (top). Helicases, such as DHX36 or DHX9, resolve the G4s and facilitate translation of the main ORF (bottom). **b** | The 80S ribosomes engaged in translation elongation stall 6–7 nucleotides prior to a G4 within the ORF. Stalling can cause ribosome backtracking and synthesis of an alternative peptide. **c** | Recognition of RNA G4s by spliceosome-associated RNA-binding proteins directs splicing of nearby introns, for example the second intron of p53.

suggested a potential function of RNA G4s in the formation of stress granules; for example, RNA G4-binding proteins, such as DHX36 and DDX3X, associate with stress granules<sup>173,174</sup>. Moreover, the C9ORF72 mRNA, which contains G4-forming repeats, the extension of which causes amyotrophic lateral sclerosis and frontotemporal dementia, promotes phase separation of stress granule proteins and granule assembly<sup>175</sup>. The propensity of mRNAs to localize to stress granules correlates with longer UTRs and coding regions and with poor translation, which are features of G4-containing mRNAs<sup>176</sup>. Finally, translation interfering tRNAs (tiRNAs) are tRNA fragments formed in stress conditions that may have roles in cancer progression<sup>177</sup>. A G4 structure in tiRNA<sup>ala</sup> or tetramolecular G4s (formed from four individual tiRNAs) appear to mediate stress granule formation<sup>178,179</sup> and translation inhibition by interacting with Y-box binding protein 1 and, subsequently, displacing eIF4F from mRNAs<sup>180</sup>. DNA analogues of tiRNA<sup>ala</sup> may trigger a neuroprotective response in motor neurons, suggesting new possibilities for interventions in neurodegenerative diseases<sup>180</sup>.

RNA G4-mediated recruitment of splicing-associated RBPs, such as heterogeneous ribonucleoprotein H (hnRNPH)<sup>181</sup> and hnRNPF<sup>182</sup>, regulates alternative splicing. In the context of its native sequence, the FMRP mRNA G4 is a potent splicing enhancer<sup>147,183</sup>. A G4 in the third intron of the p53 mRNA promotes the splicing of intron 2, and mutating the guanines of this G4 increases intron retention and synthesis of the truncated

protein  $\Delta 40p53$  (REF.<sup>184</sup>) (FIG. 5c). Likewise, a G4 promotes the correct splicing of intron 1 of the PAX9 mRNA<sup>185</sup> and another promotes the correct splicing of BACE-1 mRNA<sup>186</sup>. Recently, an RNA G4-binding ligand was found to cause thousands of alternative splicing events in cells<sup>187</sup>. Additional transcriptomics studies are required to elucidate the rule set by which G4s regulate splicing. RNA G4s are also involved in mRNA polyadenylation, piRNA biogenesis and form in ribosomal RNA (reviewed in REF.<sup>188</sup>).

### Nucleic acid and histone modifications

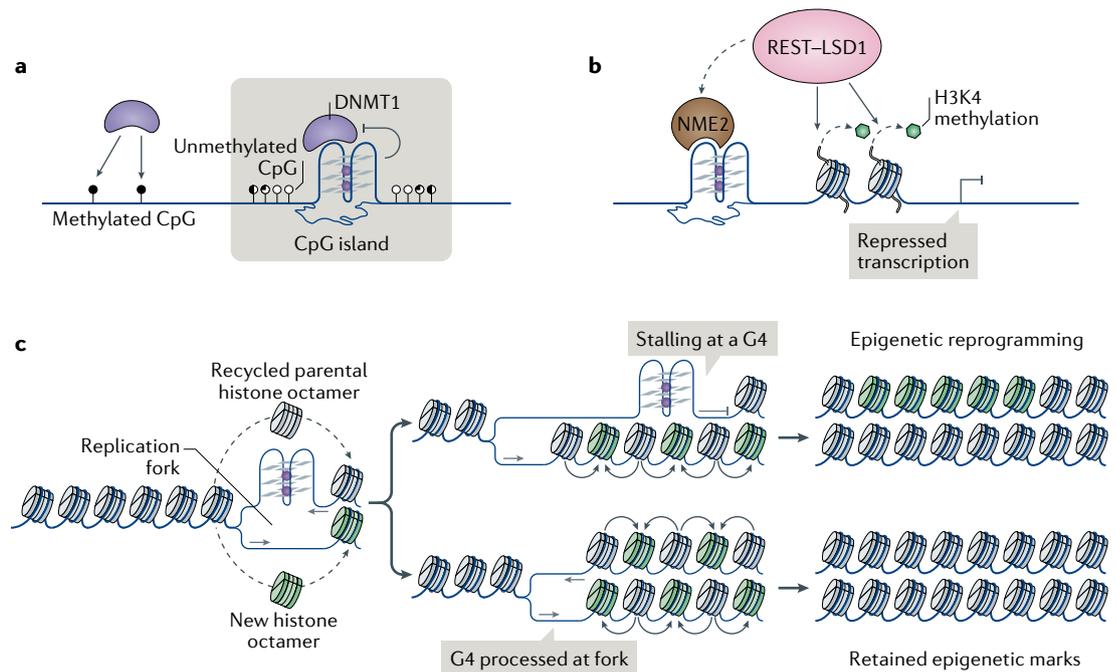
Emerging observations link G4s to covalent chemical modifications of DNA and of histones. DNA methyltransferases, which catalyse the formation of 5-methylcytosine, principally at CpG dinucleotides in mammalian cells, have a biophysical preference to bind G4 DNA over double-stranded DNA<sup>189,190</sup> (FIG. 6a). G4 binding inhibits the activity of DNA (cytosine-5)-methyltransferase 1 (DNMT1), and DNMT1 binding sites in chromatin that are marked by G4 structures are strongly hypomethylated in human leukaemia cells, prompting the hypothesis that DNMT1 is sequestered at G4 sites to inhibit methylation of proximal CpG island promoters<sup>190</sup> (FIG. 6a). In addition to DNA methylation, a recent study reported G4-dependent transcription repression of *hTERT* through non-metastatic 2 (NME2)-dependent recruitment to the promoter G4s of the RE1-silencing transcription factor (REST)-lysine-specific histone demethylase 1A (LSD1); also

#### Stress granules

Cytoplasmic membraneless bodies of proteins and RNAs that appear in response to conditions of cellular stress.

#### CpG island

A genomic region with CG:GC content higher than 60%.



**Fig. 6 | The involvement of G-quadruplexes in epigenetic control.** **a** | Binding to G-quadruplexes (G4s) inactivates DNA (cytosine-5)-methyltransferase 1 (DNMT1), thereby contributing to hypomethylation at CpG islands. **b** | Promoter G4s and their associated proteins, such as non-metastatic 2 (NME2), recruit the RE1-silencing transcription factor (REST)-lysine-specific histone demethylase 1A (LSD1) repressor complex to remove the gene-activating methylation of histone H3 Lys4 (H3K4) and repress gene expression. **c** | Stalling of DNA replication forks at G4s (for example, owing to impaired activity of helicases or the presence of G4-stabilizing ligands) may impair histone recycling through the formation of a post-replicative gap (top). Parental histones with their established modifications (grey) are lost and replaced with new histones with no or with different modifications (green), resulting in local epigenetic reprogramming.

known as KDM1A) repressor complex, which removes the gene-activating monomethylation and dimethylation of histone H3 Lys4 (H3K4me1 and H3K4me2, respectively)<sup>191</sup> (FIG. 6b). A similar mechanism was postulated for the promoter of cyclin-dependent kinase inhibitor 1 (which encodes p21<sup>Cip1</sup>), where a TRF2-G4 interaction is required to mediate REST-LSD1 activity<sup>192</sup>.

During replication, DNA synthesis and histone recycling (into the newly formed sister chromatids) are closely coordinated to ensure the maintenance of parental histone modifications in the daughter cells. In DT40 chicken cells, impaired activity of proteins required for replication of G4-forming sequences<sup>193,194</sup>, depletion of nucleotide pools (causing replication stress)<sup>195</sup> or G4 stabilization by small molecules<sup>196</sup> resulted in local alteration of epigenetic marks, including histone modifications and cytosine methylation (FIG. 6c). These observations suggest that replication fork pausing at G4 sites in conditions of replication stress can uncouple the replication machinery from histone recycling. Notably, loss of H3K4me3 at a defined G4 site led to proximal DNA cytosine methylation and heritable inactivation of BU-1 gene expression<sup>196</sup>.

RNA G4s can influence gene expression in multiple ways. For example, the formation of a G4 instead of the canonical stem-loop in certain precursor microRNAs can inhibit their maturation by Dicer<sup>31,197</sup>. Production of mature microRNAs can therefore be influenced by RNA G4-stabilizing factors, such as high K<sup>+</sup> or

Mg<sup>2+</sup> levels, or N-7 methylation of guanines<sup>198,199</sup>. Additionally, G4 formation in a mature microRNA or in its target mRNA sequence can alter the regulation of the target mRNA<sup>200,201</sup>. Cross-talk between RNA G4s and chromatin modification is exemplified by the binding of RNA G4s by Polycomb repressive complex 2 (PRC2), which catalyses the gene-repressive histone modification H3K27me3 (REF.<sup>202</sup>). An association of N<sup>6</sup>-methyladenosine (m<sup>6</sup>A) in viral RNA with predicted RNA G4s has suggested that RNA G4s may also function to guide RNA base modifications<sup>203</sup>. Multiple retrotransposons in the human genome harbour G4s, which may promote their transposition, as seen in the case of long interspersed element 1, although the underlying mechanism remains unclear<sup>204</sup>.

#### Higher-order chromatin architecture

As DNA G4 structures form in chromatin<sup>46</sup>, they might have a role in regulating higher-order genome architecture, for example in mediating promoter-enhancer interactions<sup>205</sup>. G4s may be involved in orchestrating long-range interactions, for example G4 encoding sequences that are split over long distances may come together to form a G4. In fission yeast, binding by the Rif1 protein to such half G4s is proposed to create local chromosomal compartments by enabling chromatin looping at the nuclear lamina, which may be involved in suppression of DNA replication origin firing in late S phase<sup>52</sup>. A recent computational study using a ChIP-seq

dataset from human cells indicates that DNA G4s help define higher-hierarchy chromatin domains called topological associated domains<sup>116</sup>. These studies are suggestive, and more experimentation and direct evidence is required to determine whether DNA G4s are involved in active chromatin looping.

### The therapeutic relevance of G4s

The multiple functions of G4s in DNA and in RNA collectively present opportunities for interference by small molecule-mediated manipulation of folded G4s. In this section we highlight some examples, primarily from cancer, where considerable progress towards demonstrating therapeutic potential has been reported.

### Targeting G4s

Established modes of targeting DNA with small molecules, for example through intercalation or covalent modification, have led to the development of therapeutic agents against pathogens and cancers<sup>206</sup>. Although many of these drugs are used effectively, including as first-line therapies, their use is limited owing to their toxicity and side effects. G4s could offer a new modality for targeting DNA. The distinct molecular features of G4s, in particular the G-tetrads and loops, enable structure-selective recognition by small molecules<sup>207</sup>. The functional links between G4s and gene regulation (particularly regulation of cancer genes), DNA replication and genome instability and telomere biology have prompted exploration of G4-targeting therapeutics. Initial efforts focused on targeting G4 structures at telomeres with a view to inhibiting telomere extension by telomerase in cancer cells<sup>106,208</sup>. Subsequently, it emerged that G4-binding molecules can cause DNA damage at telomeres and also at G4s throughout the genome<sup>43,209</sup>. Targeting of G4s in genes by PDS can inhibit gene expression, including numerous important oncogenes<sup>43</sup>. The naphthalene diimide G4-targeting ligand CM03 has shown promising activity against cancer cell lines and in a mouse xenograft model of pancreatic ductal adenocarcinoma, including a notable reduction in the expression of many G4-rich genes implicated in vital pathways of cancer-cell survival, metastasis and drug resistance<sup>114</sup>. Whereas early studies focused on modulating individual cancer genes by targeting their G4s<sup>111–113</sup>, the prevalence of G4s in many cancer-promoting genes suggests that collectively targeting multiple G4s, and thus inhibiting the expression of many such genes, as exemplified by CM03, would be a feasible strategy<sup>114</sup>. The observed increase in G4s in the chromatin of cancer tissues<sup>210</sup> and cell line models<sup>34,46</sup>, in comparison with normal cells and tissues, also favours targeting G4s as a general anticancer strategy.

### Synthetic lethality

The capacity of G4 ligands to specifically create synthetic lethality in tumour cells provides another potential G4-based therapeutic avenue. Synthetic lethality refers to the cell-lethal combination of two or more non-lethal genetic perturbations. This can also be mimicked chemically by pharmacological inhibition of key genes that phenocopy genetic sensitivities<sup>211</sup>. G4 ligands enhance killing of *BRCA1*-deficient or *BRCA2*-deficient cancer cells by exploiting their deficiency in homologous-directed DNA repair<sup>11,212,213</sup>. A recent genomic RNAi screen has expanded the cancer genotypes and pathways sensitive to G4-ligand treatment<sup>141</sup>. Thus, there is scope for exploring genotype-specific G4-targeting strategies, as exemplified by the G4-targeting clinical compound CX-5461, which has recently entered clinical trials for BRCA-deficient tumours<sup>41</sup> (see NCT02719977 at ClinicalTrials.gov). G4 ligands can be also used in combination with other agents. For example, the cytotoxic activities of PDS synergize with the compound NU7441, which inhibits the essential non-homologous end-joining factor DNA-dependent protein kinase<sup>213</sup>; with MK1775, which inhibits the cell cycle kinase WEE1; and with pimozone, which inhibits the deubiquitylating enzyme USP1 (also known as ubiquitin carboxyl-terminal hydrolase 1)<sup>141</sup>. These findings demonstrate the potential of G4 ligands as therapeutic agents in multiple cancer types.

### Conclusions

The study of G4s originated from curiosity-driven structural investigations, and has progressed to the point where G4s should now be considered a fundamental feature of the genome. G4s are implicated in numerous important cellular processes, in particular transcription, but also in translation and maintenance of genome stability. The key future challenge is to elucidate the details of how G4 formation is regulated, especially at gene promoters and UTRs, and the specific mechanisms underlying their biological roles. Specifically, the many G4–protein interactions that have been revealed need to be characterized in greater detail to generate a robust molecular understanding of how G4s influence protein function. Such insight will naturally lead to clearer understanding of the role G4s in disease and could ultimately be exploited in a clinical context. The study of G4s has progressed considerably over the past decade and, consequently, a framework of computational reference data and experimental tools and methodologies now exists to help drive the elucidation of the functions of G4s in finer detail over the coming years.

Published online 20 April 2020

- Bang, I. Untersuchungen über die Guanynsäure. *Biochemische* **26**, 293–311 (1910).
- Gellert, M., Lipsett, M. N. & Davies, D. R. Helix formation by guanylic acid. *Proc. Natl Acad. Sci. USA* **48**, 2013–2018 (1962).
- Sen, D. & Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334**, 364–366 (1988).  
**This paper is an early demonstration of a G4 comprising stacked tetrads with interconnecting loop sequences performed using chemical mapping and providing biological insight.**
- Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **34**, 5402–5415 (2006).
- Kwok, C. K. & Merrick, C. J. G-Quadruplexes: prediction, characterization, and biological application. *Trends Biotechnol.* **35**, 997–1013 (2017).
- Lane, A. N., Chaires, J. B., Gray, R. D. & Trent, J. O. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* **36**, 5482–5515 (2008).
- Lerner, L. K. & Sale, J. E. Replication of G Quadruplex DNA. *Genes* **10**, 95 (2019).
- Maizels, N. G4-associated human diseases. *EMBO Rep.* **16**, 910–922 (2015).
- Neidle, S. Quadruplex nucleic acids as novel therapeutic targets. *J. Med. Chem.* **59**, 5987–6011 (2016).
- Mergny, J. L. & Lacroix, L. UV melting of G-quadruplexes. *Curr. Protoc. Nucleic Acid Chem.* **37**, 17.1.1–17.1.15 (2009).
- Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33**, 2908–2916 (2005).  
**This work presents the earliest computational predictions showing that sequences encoding G4s are widespread in the human genome.**

12. Todd, A. K., Johnson, M. & Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **33**, 2901–2907 (2005).
13. Huppert, J. L. & Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **35**, 406–413 (2007).
14. Eddy, J. & Maizels, N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* **34**, 3887–3896 (2006).
15. Varizhuk, A. et al. The expanding repertoire of G4 DNA structures. *Biochimie* **135**, 54–62 (2017).
16. Stegle, O., Payet, L., Mergny, J. L., MacKay, D. J. & Leon, J. H. Predicting and understanding the stability of G-quadruplexes. *Bioinformatics* **25**, i374–i382 (2009).
17. Bedrat, A., Lacroix, L. & Mergny, J. L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **44**, 1746–1759 (2016).
18. Belmonte-Reche, E. & Morales, J. C. G4-iM Grinder: when size and frequency matter. G-Quadruplex, i-Motif and higher order structure search and analysis tool. *NAR Genom. Bioinform.* **2**, 1–12 (2020).
19. Sahakyan, A. B. et al. Machine learning model for sequence-driven DNA G-quadruplex formation. *Sci. Rep.* **7**, 14535 (2017).
20. Garant, J. M., Perreault, J. P. & Scott, M. S. Motif independent identification of potential RNA G-quadruplexes by G4RNA screener. *Bioinformatics* **33**, 3532–3537 (2017).
21. Woodford, K. J., Howell, R. M. & Udsin, K. A novel K<sup>+</sup>-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes. *J. Biol. Chem.* **269**, 27029–27035 (1994).
22. Han, H., Hurley, L. H. & Salazar, M. A DNA polymerase stop assay for G-quadruplex-interactive compounds. *Nucleic Acids Res.* **27**, 537–542 (1999).
23. Kwok, C. K. & Balasubramanian, S. Targeted detection of G-quadruplexes in cellular RNAs. *Angew. Chem. Int. Ed. Engl.* **54**, 6751–6754 (2015).
24. Chambers, V. S. et al. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* **33**, 877–881 (2015).
25. Marsico, G. et al. Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res.* **47**, 3862–3874 (2019).
26. Guo, J. U. & Bartel, D. P. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **353**, aaf5371 (2016).
27. Kwok, C. K., Marsico, G., Sahakyan, A. B., Chambers, V. S. & Balasubramanian, S. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Methods* **13**, 841–844 (2016).
28. Kouzine, F. et al. Permanganate/S1 nuclease footprinting reveals non-B DNA structures with regulatory potential across a mammalian genome. *Cell Syst.* **4**, 344–356 (2017).
29. Williamson, J. R., Raghuraman, M. K. & Cech, T. R. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell* **59**, 871–880 (1989).
30. Wilkinson, K. A., Merino, E. J. & Weeks, K. M. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.* **1**, 1610–1616 (2006).
31. Kwok, C. K., Sahakyan, A. B. & Balasubramanian, S. Structural analysis using SHALiPE to reveal RNA G-quadruplex formation in human precursor microRNA. *Angew. Chem. Int. Ed. Engl.* **55**, 8958–8961 (2016).
32. Kwok, C. K., Marsico, G. & Balasubramanian, S. Detecting RNA G-quadruplexes (rG4s) in the transcriptome. *Cold Spring Harb. Perspect. Biol.* **10**, a032284 (2018).
33. Schaffitzel, C. et al. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylomychia lemnae* macronuclei. *Proc. Natl Acad. Sci. USA* **98**, 8572–8577 (2001).
34. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **5**, 182–186 (2013).  
**This work is the first demonstration of G4s in human cells by imaging using a structure-specific antibody.**
35. Henderson, A. et al. Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.* **42**, 860–869 (2014).
36. Liu, H. Y. et al. Conformation selective antibody enables genome profiling and leads to discovery of parallel G-quadruplex in human telomeres. *Cell Chem. Biol.* **23**, 1261–1270 (2016).
37. Biffi, G., Di Antonio, M., Tannahill, D. & Balasubramanian, S. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.* **6**, 75–80 (2014).
38. Wang, Y. et al. G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma. *Nat. Commun.* **10**, 943 (2019).
39. Zhang, M. et al. Mammalian CST averts replication failure by preventing G-quadruplex accumulation. *Nucleic Acids Res.* **47**, 5243–5259 (2019).
40. Wu, W. et al. HERC2 facilitates BLM and WRN helicase complex interaction with RPA to suppress G-quadruplex DNA. *Cancer Res.* **78**, 6371–6385 (2018).
41. Xu, H. et al. CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nat. Commun.* **8**, 14432 (2017).
42. Kazemier, H. G., Paeschke, K. & Lansdorf, P. M. Guanidine quadruplex monoclonal antibody 1H6 cross-reacts with restrained thymidine-rich single stranded DNA. *Nucleic Acids Res.* **45**, 5913–5919 (2017).
43. Rodriguez, R. et al. Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.* **8**, 301–310 (2012).
44. Lefebvre, J., Guetta, C., Poyer, F., Mahuteau-Betzer, F. & Teulade-Fichou, M. P. Copper–alkyne complexation responsible for the nucleolar localization of quadruplex nucleic acid drugs labeled by click reactions. *Angew. Chem. Int. Ed. Engl.* **56**, 11365–11369 (2017).
45. Shivalingam, A. et al. The interactions between a small molecule and G-quadruplexes are visualized by fluorescence lifetime imaging microscopy. *Nat. Commun.* **6**, 8178 (2015).
46. Hansel-Hertsch, R. et al. G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **48**, 1267–1272 (2016).  
**This paper presents the first maps of G4s generated in an endogenous chromatin context using G4 ChIP-seq and demonstrating G4 enrichment in active promoters linked with elevated transcription.**
47. Hansel-Hertsch, R., Spiegel, J., Marsico, G., Tannahill, D. & Balasubramanian, S. Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nat. Protoc.* **13**, 551–564 (2018).
48. 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).
49. Gray, L. T., Vallur, A. C., Eddy, J. & Maizels, N. G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. *Nat. Chem. Biol.* **10**, 313–318 (2014).
50. Paeschke, K., Capra, J. A. & Zakian, V. A. DNA replication through G-quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. *Cell* **145**, 678–691 (2011).  
**This paper presents genetic experiments in yeast showing that the helicase Pif1 resolves G4s in vivo to prevent replication-fork stalling and DNA breaks.**
51. Gotz, S., Pandey, S., Bartsch, S., Juranek, S. & Paeschke, K. A novel G-quadruplex binding protein in yeast-Slx9. *Molecules* **24**, 1774 (2019).
52. Kanoh, Y. et al. Rif1 binds to G quadruplexes and suppresses replication over long distances. *Nat. Struct. Mol. Biol.* **22**, 889–897 (2015).
53. Meyer, C. A. & Liu, X. S. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nat. Rev. Genet.* **15**, 709–721 (2014).
54. Herdy, B. et al. Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts. *Nucleic Acids Res.* **46**, 11592–11604 (2018).
55. Murat, P. et al. RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. *Genome Biol.* **19**, 229 (2018).
56. Sauer, M. et al. DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions. *Nat. Commun.* **10**, 2421 (2019).
57. Joachimi, A., Benz, A. & Hartig, J. S. A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorg. Med. Chem.* **17**, 6811–6815 (2009).
58. Bugaut, A. & Balasubramanian, S. A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry* **47**, 689–697 (2008).
59. Hazel, P., Huppert, J., Balasubramanian, S. & Neidle, S. Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.* **126**, 16405–16415 (2004).
60. Sen, D. & Gilbert, W. A sodium–potassium switch in the formation of four-stranded G4-DNA. *Nature* **344**, 410–414 (1990).
61. Zoroddu, M. A. et al. The essential metals for humans: a brief overview. *J. Inorg. Biochem.* **195**, 120–129 (2019).
62. Selvam, S., Koirala, D., Yu, Z. & Mao, H. Quantification of topological coupling between DNA superhelicity and G-quadruplex formation. *J. Am. Chem. Soc.* **136**, 13967–13970 (2014).
63. Shrestha, P. et al. Confined space facilitates G-quadruplex formation. *Nat. Nanotechnol.* **12**, 582–588 (2017).
64. Fry, M. & Loeb, L. A. Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CCG)n. *J. Biol. Chem.* **274**, 12797–12802 (1999).
65. Mohaghegh, P., Karow, J. K., Brosh, R. M. Jr, Bohr, V. A. & Hickson, I. D. The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res.* **29**, 2843–2849 (2001).
66. Creacy, S. D. et al. G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates. *J. Biol. Chem.* **283**, 34626–34634 (2008).
67. Chatterjee, S. et al. Mechanistic insight into the interaction of BLM helicase with intra-strand G-quadruplex structures. *Nat. Commun.* **5**, 5556 (2014).
68. Chen, M. C. et al. Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature* **558**, 465–469 (2018).  
**This work presents the first X-ray crystallography structure of a G4-resolving helicase bound to a G4, proposing a structural mechanism for G4 unfolding.**
69. Tippiana, R., Hwang, H., Opreko, P. L., Bohr, V. A. & Myong, S. Single-molecule imaging reveals a common mechanism shared by G-quadruplex-resolving helicases. *Proc. Natl Acad. Sci. USA* **113**, 8448–8453 (2016).
70. Tippiana, R., Chen, M. C., Demeshkina, N. A., Ferre-D'Amare, A. R. & Myong, S. RNA G-quadruplex is resolved by repetitive and ATP-dependent mechanism of DHX36. *Nat. Commun.* **10**, 1855 (2019).
71. Byrd, A. K. & Raney, K. D. Structure and function of Pif1 helicase. *Biochem. Soc. Trans.* **45**, 1159–1171 (2017).
72. Wu, Y., Shin-ya, K. & Brosh, R. M. Jr FANCD1 helicase defective in Fanconi anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Mol. Cell Biol.* **28**, 4116–4128 (2008).
73. Benhalevy, D. et al. The human CCHC-type zinc finger nucleic acid-binding protein binds G-rich elements in target mRNA coding sequences and promotes translation. *Cell Rep.* **18**, 2979–2990 (2017).
74. Pietras, Z. et al. Dedicated surveillance mechanism controls G-quadruplex forming non-coding RNAs in human mitochondria. *Nat. Commun.* **9**, 2558 (2018).
75. Ray, S., Bandaria, J. N., Qureshi, M. H., Yildiz, A. & Balci, H. G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. *Proc. Natl Acad. Sci. USA* **111**, 2990–2995 (2014).
76. Zaug, A. J., Podell, E. R. & Cech, T. R. Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc. Natl Acad. Sci. USA* **102**, 10864–10869 (2005).
77. Gonzalez, V., Guo, K., Hurley, L. & Sun, D. Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J. Biol. Chem.* **284**, 23622–23635 (2009).
78. Niu, K. et al. Identification of LARK as a novel and conserved G-quadruplex binding protein in invertebrates and vertebrates. *Nucleic Acids Res.* **47**, 7306–7320 (2019).
79. Serikawa, T. et al. Comprehensive identification of proteins binding to RNA G-quadruplex motifs in the 5' UTR of tumor-associated mRNAs. *Biochimie* **144**, 169–184 (2018).
80. Kouzine, F., Liu, J., Sanford, S., Chung, H. J. & Levens, D. The dynamic response of upstream DNA to transcription-generated torsional stress. *Nat. Struct. Mol. Biol.* **11**, 1092–1100 (2004).
81. Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F. & Maizels, N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures

- containing G4 DNA. *Genes. Dev.* **18**, 1618–1629 (2004).
82. Chen, L. et al. R-ChIP using inactive RNase H reveals dynamic coupling of R-loops with transcriptional pausing at gene promoters. *Mol. Cell* **68**, 745–757 (2017).
  83. Gehring, K., Leroy, J. L. & Gueron, M. A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature* **363**, 561–565 (1993).
  84. Zeraati, M. et al. I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.* **10**, 631–637 (2018).
  85. Cui, Y., Kong, D., Ghimire, C., Xu, C. & Mao, H. Mutually exclusive formation of G-quadruplex and i-motif is a general phenomenon governed by steric hindrance in duplex DNA. *Biochemistry* **55**, 2291–2299 (2016).
  86. Hoffmann, R. F. et al. Guanine quadruplex structures localize to heterochromatin. *Nucleic Acids Res.* **44**, 152–163 (2016).
  87. Sen, D. & Poon, L. C. RNA and DNA complexes with hemin [Fe(III) heme] are efficient peroxidases and peroxygenases: how do they do it and what does it mean? *Crit. Rev. Biochem. Mol. Biol.* **46**, 478–492 (2011).
  88. Gray, L. T. et al. G-quadruplexes sequester free heme in living cells. *Cell Chem. Biol.* **26**, 1681–1691 (2019).
  89. Fouquerel, E. et al. Targeted and persistent 8-oxoguanine base damage at telomeres promotes telomere loss and crisis. *Mol. Cell* **75**, 117–130 (2019).
  90. Fleming, A. M., Zhu, J., Ding, Y. & Burrows, C. J. 8-Oxo-7,8-dihydroguanine in the context of a gene promoter G-quadruplex is an on-off switch for transcription. *ACS Chem. Biol.* **12**, 2417–2426 (2017).
  91. Shay, J. W. & Wright, W. E. Telomeres and telomerase: three decades of progress. *Nat. Rev. Genet.* **20**, 299–309 (2019).
  92. Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. & Lipps, H. J. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol.* **12**, 847–854 (2005).
  93. de Lange, T. T-loops and the origin of telomeres. *Nat. Rev. Mol. Cell Biol.* **5**, 323–329 (2004).
  94. Biffi, G., Tannahill, D. & Balasubramanian, S. An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2. *J. Am. Chem. Soc.* **134**, 11974–11976 (2012).
  95. Smith, J. S. et al. Rudimentary G-quadruplex-based telomere capping in *Saccharomyces cerevisiae*. *Nat. Struct. Mol. Biol.* **18**, 478–485 (2011).
  96. Takahama, K. et al. Regulation of telomere length by G-quadruplex telomere DNA- and TERRA-binding protein TLS/FUS. *Chem. Biol.* **20**, 341–350 (2013).
  97. Takahama, K., Kino, K., Arai, S., Kurokawa, R. & Oyoshi, T. Identification of Ewing's sarcoma protein as a G-quadruplex DNA- and RNA-binding protein. *FEBS J.* **278**, 988–998 (2011).
  98. Vannier, J. B., Pavic-Kaltenbrunner, V., Petalcorin, M. I., Ding, H. & Boulton, S. J. RTF1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. *Cell* **149**, 795–806 (2012).
  99. Sfeir, A. et al. Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell* **138**, 90–103 (2009).
  100. Zahler, A. M., Williamson, J. R., Cech, T. R. & Prescott, D. M. Inhibition of telomerase by G-quartet DNA structures. *Nature* **350**, 718–720 (1991). **This paper is the earliest demonstration that the activity of telomerase can be affected by a G4 structure in its telomere DNA substrate, suggesting that G4s might regulate telomere elongation.**
  101. Moye, A. L. et al. Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.* **6**, 7643 (2015).
  102. Zhang, M. L. et al. Yeast telomerase subunit Est1p has guanine quadruplex-promoting activity that is required for telomere elongation. *Nat. Struct. Mol. Biol.* **17**, 202–209 (2010).
  103. Hwang, H., Buncher, N., Opreks, P. L. & Myong, S. POT1–TPP1 regulates telomeric overhang structural dynamics. *Structure* **20**, 1872–1880 (2012).
  104. Jansson, L. I. et al. Telomere DNA G-quadruplex folding within actively extending human telomerase. *Proc. Natl Acad. Sci. USA* **116**, 9350–9359 (2019).
  105. Booy, E. P. et al. The RNA helicase RHAU (DHX36) unwinds a G4-quadruplex in human telomerase RNA and promotes the formation of the P1 helix template boundary. *Nucleic Acids Res.* **40**, 4110–4124 (2012).
  106. Sun, D. et al. Inhibition of human telomerase by a G-quadruplex-interactive compound. *J. Med. Chem.* **40**, 2113–2116 (1997).
  107. Neidle, S. Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J.* **277**, 1118–1125 (2010).
  108. Clynes, D. et al. Suppression of the alternative lengthening of telomere pathway by the chromatin remodelling factor ATRX. *Nat. Commun.* **6**, 7538 (2015).
  109. Gowan, S. M., Heald, R., Stevens, M. F. & Kelland, L. R. Potent inhibition of telomerase by small-molecule pentacyclic acridines capable of interacting with G-quadruplexes. *Mol. Pharmacol.* **60**, 981–988 (2001).
  110. Simonsson, T., Pecinka, P. & Kubista, M. DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.* **26**, 1167–1172 (1998).
  111. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J. & Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl Acad. Sci. USA* **99**, 11593–11598 (2002). **This study of the G4 structure in the MYC promoter shows that a small-molecule G4 ligand can inhibit transcription.**
  112. Cogo, S. & Xodo, L. E. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.* **34**, 2536–2549 (2006).
  113. Bejugam, M. et al. Trisubstituted isoalloxazines as a new class of G-quadruplex binding ligands: small molecule regulation of c-kit oncogene expression. *J. Am. Chem. Soc.* **129**, 12926–12927 (2007).
  114. Marchetti, C. et al. Targeting multiple effector pathways in pancreatic ductal adenocarcinoma with a G-quadruplex-binding small molecule. *J. Med. Chem.* **61**, 2500–2517 (2018).
  115. Kumar, P. et al. Zinc-finger transcription factors are associated with guanine quadruplex motifs in human, chimpanzee, mouse and rat promoters genome-wide. *Nucleic Acids Res.* **39**, 8005–8016 (2011).
  116. Hou, Y. et al. Integrative characterization of G-Quadruplexes in the three-dimensional chromatin structure. *Epigenetics* **14**, 894–911 (2019).
  117. Thakur, R. K. et al. Metastases suppressor NM23-H2 interaction with G-quadruplex DNA within c-MYC promoter nuclease hypersensitive element induces c-MYC expression. *Nucleic Acids Res.* **37**, 172–183 (2009).
  118. Borgognone, M., Armas, P. & Calcaterra, N. B. Cellular nucleic-acid-binding protein, a transcriptional enhancer of c-Myc, promotes the formation of parallel G-quadruplexes. *Biochem. J.* **428**, 491–498 (2010).
  119. Raiber, E. A., Kranaster, R., Lam, E., Nikan, M. & Balasubramanian, S. A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Res.* **40**, 1499–1508 (2012).
  120. Li, P. T. et al. Expression of the human telomerase reverse transcriptase gene is modulated by quadruplex formation in its first exon due to DNA methylation. *J. Biol. Chem.* **292**, 20859–20870 (2017).
  121. Agarwal, T., Roy, S., Kumar, S., Chakraborty, T. K. & Maiti, S. In the sense of transcription regulation by G-quadruplexes: asymmetric effects in sense and antisense strands. *Biochemistry* **53**, 3711–3718 (2014).
  122. Holder, I. T. & Hartig, J. S. A matter of location: influence of G-quadruplexes on *Escherichia coli* gene expression. *Chem. Biol.* **21**, 1511–1521 (2014).
  123. Belotserkovskii, B. P., Soo Shin, J. H. & Hanawalt, P. C. Strong transcription blockage mediated by R-loop formation within a G-rich homopurine-homopyrimidine sequence localized in the vicinity of the promoter. *Nucleic Acids Res.* **45**, 6589–6599 (2017).
  124. Eddy, J. et al. G4 motifs correlate with promoter-proximal transcriptional pausing in human genes. *Nucleic Acids Res.* **39**, 4975–4983 (2011).
  125. Du, Z., Zhao, Y. & Li, N. Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. *Genome Res.* **18**, 233–241 (2008).
  126. Wanrooij, P. H. et al. A hybrid G-quadruplex structure formed between RNA and DNA explains the extraordinary stability of the mitochondrial R-loop. *Nucleic Acids Res.* **40**, 10334–10344 (2012).
  127. Zheng, K. W. et al. Co-transcriptional formation of DNA:RNA hybrid G-quadruplex and potential function as constitutively cis element for transcription control. *Nucleic Acids Res.* **41**, 5533–5541 (2013).
  128. Puget, N., Miller, K. M. & Legube, G. Non-canonical DNA/RNA structures during transcription-coupled double-strand break repair: roadblocks or bona fide repair intermediates? *DNA Repair* **81**, 102661 (2019).
  129. Techer, H., Koundrioukoff, S., Nicolas, A. & Debatisse, M. The impact of replication stress on replication dynamics and DNA damage in vertebrate cells. *Nat. Rev. Genet.* **18**, 535–550 (2017).
  130. De, S. & Michor, F. DNA secondary structures and epigenetic determinants of cancer genome evolution. *Nat. Struct. Mol. Biol.* **18**, 950–955 (2011).
  131. Georgakopoulos-Soares, I., Morganella, S., Jain, N., Hemberg, M. & Nik-Zainal, S. Noncanonical secondary structures arising from non-B DNA motifs are determinants of mutagenesis. *Genome Res.* **28**, 1264–1271 (2018).
  132. Kruisselbrink, E. et al. Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCD1-defective *C. elegans*. *Curr. Biol.* **18**, 900–905 (2008).
  133. Castillo Bosch, P. et al. FANCD1 promotes DNA synthesis through G-quadruplex structures. *EMBO J.* **33**, 2521–2533 (2014).
  134. Lemmens, B., van Schendel, R. & Tijsterman, M. Mutagenic consequences of a single G-quadruplex demonstrate mitotic inheritance of DNA replication fork barriers. *Nat. Commun.* **6**, 8909 (2015).
  135. Paeschke, K. et al. Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature* **497**, 458–462 (2013).
  136. Piazza, A. et al. Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **38**, 4337–4348 (2010).
  137. Piazza, A. et al. Non-canonical G-quadruplexes cause the hCEB1 minisatellite instability in *Saccharomyces cerevisiae*. *eLife* **6**, e26884 (2017).
  138. Piazza, A. et al. Short loop length and high thermal stability determine genomic instability induced by G-quadruplex-forming minisatellites. *EMBO J.* **34**, 1718–1734 (2015).
  139. van Wietmarschen, N. et al. BLM helicase suppresses recombination at G-quadruplex motifs in transcribed genes. *Nat. Commun.* **9**, 271 (2018).
  140. Pladevall-Morerà, D. et al. Proteomic characterization of chromosomal common fragile site (CFS)-associated proteins uncovers ATRX as a regulator of CFS stability. *Nucleic Acids Res.* **47**, 8004–8018 (2019).
  141. Zyner, K. G. et al. Genetic interactions of G-quadruplexes in humans. *eLife* **8**, e46793 (2019).
  142. Muller, S. et al. Pyridostatin analogues promote telomere dysfunction and long-term growth inhibition in human cancer cells. *Org. Biomol. Chem.* **10**, 6537–6546 (2012).
  143. 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).
  144. Ginno, P. A., Lim, Y. W., Lott, P. L., Korf, I. & Chedin, 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).
  145. De Magis, A. et al. DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proc. Natl Acad. Sci. USA* **116**, 816–825 (2019).
  146. Nguyen, D. T. et al. The chromatin remodelling factor ATRX suppresses R-loops in transcribed telomeric repeats. *EMBO Rep.* **18**, 914–928 (2017).
  147. Schaeffer, C. et al. The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J.* **20**, 4803–4813 (2001).
  148. Kumari, S., Bugaut, A., Huppert, J. L. & Balasubramanian, S. An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.* **3**, 218–221 (2007).
  149. Shahid, R., Bugaut, A. & Balasubramanian, S. The BCL-2 5' untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. *Biochemistry* **49**, 8300–8306 (2010).
  150. Lammich, S. et al. Translational repression of the disintegrin and metalloprotease ADAM10 by a stable G-quadruplex secondary structure in its 5'-untranslated region. *J. Biol. Chem.* **286**, 45063–45072 (2011).
  151. Khateb, S., Weisman-Shomer, P., Hershco-Shani, I., Ludvig, A. L. & Fry, M. D. The tetraplex (CGG)<sub>n</sub> destabilizing proteins hnRNP A2 and CBF-A enhance the in vivo translation of fragile X premutation mRNA. *Nucleic Acids Res.* **35**, 5775–5788 (2007).

152. Kumari, S., Bugaut, A. & Balasubramanian, S. Position and stability are determining factors for translation repression by an RNA G-quadruplex-forming sequence within the 5' UTR of the NRAS proto-oncogene. *Biochemistry* **47**, 12664–12669 (2008).
153. Huppert, J. L., Bugaut, A., Kumari, S. & Balasubramanian, S. G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res.* **36**, 6260–6268 (2008).
154. Hinnebusch, A. G. The scanning mechanism of eukaryotic translation initiation. *Annu. Rev. Biochem.* **83**, 779–812 (2014).
155. Wolfe, A. L. et al. RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* **513**, 65–70 (2014).  
**This study presents a comprehensive investigation into the control of translation by the helicase eIF4A–RNA G4 interactions in the 5' UTR of mRNAs.**
156. Modelska, A. et al. The malignant phenotype in breast cancer is driven by eIF4A1-mediated changes in the translational landscape. *Cell Death Dis.* **6**, e1603 (2015).
157. Bonnal, S. et al. A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. *J. Biol. Chem.* **278**, 39330–39336 (2003).
158. Koukouraki, P. & Doxakis, E. Constitutive translation of human alpha-synuclein is mediated by the 5'-untranslated region. *Open Biol.* **6**, 160022 (2016).
159. Morris, M. J., Negishi, Y., Pászint, C., Schonhoft, J. D. & Basu, S. An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES. *J. Am. Chem. Soc.* **132**, 17831–17839 (2010).
160. Bhattacharyya, D., Diamond, P. & Basu, S. An independently folding RNA G-quadruplex domain directly recruits the 40S ribosomal subunit. *Biochemistry* **54**, 1879–1885 (2015).
161. Cammas, A. et al. Stabilization of the G-quadruplex at the VEGF IRES represses cap-independent translation. *RNA Biol.* **12**, 320–329 (2015).
162. Wu, Y. et al. Stabilization of VEGF G-quadruplex and inhibition of angiogenesis by quindoline derivatives. *Biochim. Biophys. Acta* **1840**, 2970–2977 (2014).
163. Endoh, T., Kawasaki, Y. & Sugimoto, N. Suppression of gene expression by G-quadruplexes in open reading frames depends on G-quadruplex stability. *Angew. Chem. Int. Ed. Engl.* **52**, 5522–5526 (2013).
164. Mirihana Arachchilage, G., Hetti Arachchilage, M., Venkataraman, A., Piontkivska, H. & Basu, S. Stable G-quadruplex enabling sequences are selected against by the context-dependent codon bias. *Gene* **696**, 149–161 (2019).
165. Hagerman, R. J. et al. Fragile X syndrome. *Nat. Rev. Dis. Prim.* **3**, 17065 (2017).
166. Agarwala, P., Pandey, S., Mapa, K. & Maiti, S. The G-quadruplex augments translation in the 5' untranslated region of transforming growth factor  $\beta$ 2. *Biochemistry* **52**, 1528–1538 (2013).
167. Serikawa, T., Eberle, J. & Kurreck, J. Effects of genomic disruption of a guanine quadruplex in the 5' UTR of the Bcl-2 mRNA in melanoma cells. *FEBS Lett.* **591**, 3649–3659 (2017).
168. Bhattacharyya, D. et al. Engineered domain swapping indicates context dependent functional role of RNA G-quadruplexes. *Biochimie* **137**, 147–150 (2017).
169. Endoh, T. & Sugimoto, N. Conformational dynamics of the RNA G-quadruplex and its effect on translation efficiency. *Molecules* **24**, 1613 (2019).
170. Bugaut, A., Murat, P. & Balasubramanian, S. An RNA hairpin to G-quadruplex conformational transition. *J. Am. Chem. Soc.* **134**, 19953–19956 (2012).
171. Subramanian, M. et al. G-quadruplex RNA structure as a signal for neurite mRNA targeting. *EMBO Rep.* **12**, 697–704 (2011).
172. Darnell, J. C. et al. Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell* **107**, 489–499 (2001).
173. Valentin-Vega, Y. A. et al. Cancer-associated DDX3X mutations drive stress granule assembly and impair global translation. *Sci. Rep.* **6**, 25996 (2016).
174. Chalupnikova, K. et al. Recruitment of the RNA helicase RHAU to stress granules via a unique RNA-binding domain. *J. Biol. Chem.* **283**, 35186–35198 (2008).
175. Fay, M. M., Anderson, P. J. & Ivanov, P. ALS/FTD-associated C9ORF72 repeat RNA promotes phase transitions in vitro and in cells. *Cell Rep.* **21**, 3573–3584 (2017).
176. Khong, A. et al. The stress granule transcriptome reveals principles of mRNA accumulation in stress granules. *Mol. Cell* **68**, 808–820 (2017).
177. Tao, E. W., Cheng, W. Y., Li, W. L., Yu, J. & Gao, Q. Y. tiRNAs: a novel class of small noncoding RNAs that helps cells respond to stressors and plays roles in cancer progression. *J. Cell Physiol.* **235**, 683–690 (2020).
178. Lyons, S. M., Achorn, C., Kedersha, N. L., Anderson, P. J. & Ivanov, P. YB-1 regulates tiRNA-induced stress granule formation but not translational repression. *Nucleic Acids Res.* **44**, 6949–6960 (2016).
179. Lyons, S. M., Gudanis, D., Coyne, S. M., Gdaniec, Z. & Ivanov, P. Identification of functional tetramolecular RNA G-quadruplexes derived from transfer RNAs. *Nat. Commun.* **8**, 1127 (2017).
180. Ivanov, P. et al. G-quadruplex structures contribute to the neuroprotective effects of angiogenin-induced tRNA fragments. *Proc. Natl Acad. Sci. USA* **111**, 18201–18206 (2014).
181. Conlon, E. G. et al. The C9ORF72 GGGGCC expansion forms RNA G-quadruplex structures and sequesters hnRNP H to disrupt splicing in ALS brains. *eLife* **5**, e17820 (2016).
182. Huang, H., Zhang, J., Harvey, S. E., Hu, X. & Cheng, C. RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNP. *Genes Dev.* **31**, 2296–2309 (2017).
183. Didiot, M. C. et al. The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. *Nucleic Acids Res.* **36**, 4902–4912 (2008).
184. Marcel, V. et al. G-quadruplex structures in TP53 intron 5: role in alternative splicing and in production of p53 mRNA isoforms. *Carcinogenesis* **32**, 271–278 (2011).
185. Ribeiro, M. M. et al. G-quadruplex formation enhances splicing efficiency of PAX9 intron 1. *Hum. Genet.* **134**, 37–44 (2015).
186. Fiset, J. F., Montagna, D. R., Mihalescu, M. R. & Wolfe, M. S. A G-rich element forms a G-quadruplex and regulates BACE1 mRNA alternative splicing. *J. Neurochem.* **121**, 763–773 (2012).
187. Zhang, J., Harvey, S. E. & Cheng, C. A high-throughput screen identifies small molecule modulators of alternative splicing by targeting RNA G-quadruplexes. *Nucleic Acids Res.* **47**, 3667–3679 (2019).
188. Kharel, P., Balaratnam, S., Beals, N. & Basu, S. The role of RNA G-quadruplexes in human diseases and therapeutic strategies. *Wiley Interdiscip. Rev. RNA* **11**, e1568 (2020).
189. Cree, S. L. et al. DNA G-quadruplexes show strong interaction with DNA methyltransferases in vitro. *FEBS Lett.* **590**, 2870–2883 (2016).
190. Mao, S. Q. et al. DNA G-quadruplex structures mold the DNA methylome. *Nat. Struct. Mol. Biol.* **25**, 951–957 (2018).
191. Saha, D. et al. Epigenetic suppression of human telomerase (hTERT) is mediated by the metastasis suppressor NME2 in a G-quadruplex-dependent fashion. *J. Biol. Chem.* **292**, 15205–15215 (2017).
192. Hussain, T. et al. Transcription regulation of CDKN1A (p21/CIP1/WAF1) by TRF2 is epigenetically controlled through the REST repressor complex. *Sci. Rep.* **7**, 11541 (2017).
193. Sarkies, P. et al. FANCI coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. *Nucleic Acids Res.* **40**, 1485–1498 (2012).
194. Sarkies, P., Reams, C., Simpson, L. J. & Sale, J. E. Epigenetic instability due to defective replication of structured DNA. *Mol. Cell* **40**, 703–713 (2010).  
**This paper presents genetic experiments that provide the first demonstration of a link between G4s and cellular epigenetic inheritance.**
195. Papadopoulos, C., Guilbaud, G., Schiavone, D. & Sale, J. E. Nucleotide pool depletion induces G-quadruplex-dependent perturbation of gene expression. *Cell Rep.* **13**, 2491–2503 (2015).
196. Guilbaud, G. et al. Local epigenetic reprogramming induced by G-quadruplex ligands. *Nat. Chem.* **9**, 1110–1117 (2017).
197. Mirihana Arachchilage, G., Dassanayake, A. C. & Basu, S. A potassium ion-dependent RNA structural switch regulates human pre-miRNA 92b maturation. *Chem. Biol.* **22**, 262–272 (2015).
198. Pandey, S., Agarwala, P., Jayaraj, G. G., Gargallo, R. & Maiti, S. The RNA stem-loop to G-quadruplex equilibrium controls mature microRNA production inside the cell. *Biochemistry* **54**, 7067–7078 (2015).
199. Pandolfini, L. et al. METTL1 promotes let-7 microRNA processing via m7G methylation. *Mol. Cell* **74**, 1278–1290.e9 (2019).
200. Chan, K. L. et al. Structural analysis reveals the formation and role of RNA G-quadruplex structures in human mature microRNAs. *Chem. Commun.* **54**, 10878–10881 (2018).
201. Rouleau, S., Glouzon, J. S., Brumwell, A., Bisailon, M. & Perreault, J. P. 3' UTR G-quadruplexes regulate miRNA binding. *RNA* **23**, 1172–1179 (2017).
202. Wang, X. et al. Targeting of polycomb repressive complex 2 to RNA by short repeats of consecutive guanines. *Mol. Cell* **65**, 1056–1067.e5 (2017).
203. Fleming, A. M., Nguyen, N. L. B. & Burrows, C. J. Colocalization of m<sup>6</sup>A and G-quadruplex-forming sequences in viral RNA (HIV, Zika, hepatitis B, and SV40) suggests topological control of adenosine N<sup>6</sup>-methylation. *ACS Cent. Sci.* **5**, 218–228 (2019).
204. Sahakyan, A. B., Murat, P., Mayer, C. & Balasubramanian, S. G-quadruplex structures within the 3' UTR of LINE-1 elements stimulate retrotransposition. *Nat. Struct. Mol. Biol.* **24**, 243–247 (2017).
205. Hegyi, H. Enhancer–promoter interaction facilitated by transiently forming G-quadruplexes. *Sci. Rep.* **5**, 9165 (2015).
206. Silverman, R. B. & Holladay, M. W. *The Organic Chemistry of Drug Design and Drug Action* 3rd edn (Academic, 2015).
207. Neidle, S. Quadruplex nucleic acids as targets for anticancer therapeutics. *Nat. Rev. Chem.* **1**, 0041 (2017).
208. Neidle, S. & Parkinson, G. Telomere maintenance as a target for anticancer drug discovery. *Nat. Rev. Drug Discov.* **1**, 383–393 (2002).
209. Salvati, E. et al. PARP1 is activated at telomeres upon G4 stabilization: possible target for telomere-based therapy. *Oncogene* **29**, 6280–6293 (2010).
210. Biffi, G., Tannahill, D., Miller, J., Howat, W. J. & Balasubramanian, S. Elevated levels of G-quadruplex formation in human stomach and liver cancer tissues. *PLOS ONE* **9**, e102711 (2014).
211. O'Neil, N. J., Bailey, M. L. & Hieter, P. Synthetic lethality and cancer. *Nat. Rev. Genet.* **18**, 613–623 (2017).
212. Zimmer, J. et al. Targeting BRCA1 and BRCA2 deficiencies with G-quadruplex-interacting compounds. *Mol. Cell* **61**, 449–460 (2016).
213. McLuckie, K. I. et al. G-quadruplex DNA as a molecular target for induced synthetic lethality in cancer cells. *J. Am. Chem. Soc.* **135**, 9640–9643 (2013).
214. Parkinson, G. N., Lee, M. P. & Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417**, 876–880 (2002).

## Acknowledgements

The Balasubramanian laboratory is supported by Cancer Research UK core and programme award funding (C14303/A17197; C9681/A18618). S.B. is a Senior Investigator of the Wellcome Trust (099232/Z/12/Z) and D.V. is a Herchel Smith postdoctoral fellow. J.S. gratefully acknowledges EU H2020 Framework Programme funding (H2020-MSCA-IF-2016, ID: 747297-QAPS).

## Author contributions

The authors contributed equally to all aspects of the article.

## Competing interests

S.B. is a founder and shareholder of Cambridge Epigenetix Ltd.

## Peer review information

*Nature Reviews Molecular Cell Biology* thanks Cyril Dominguez, Sua Myong and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

## Publisher's note

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

© Springer Nature Limited 2020

ARTICLE

<https://doi.org/10.1038/s41467-019-11104-0>

OPEN

# Involvement of G-quadruplex regions in mammalian replication origin activity

Paulina Prorok<sup>1</sup>, Marie Artufel<sup>2,10</sup>, Antoine Aze<sup>1,10</sup>, Philippe Coulombe<sup>1,10</sup>, Isabelle Peiffer<sup>1</sup>, Laurent Lacroix<sup>3</sup>, Aurore Guédin<sup>4</sup>, Jean-Louis Mergny <sup>4,5</sup>, Julia Damaschke<sup>6</sup>, Aloys Schepers<sup>6,7</sup>, Christelle Cayrou<sup>1,9</sup>, Marie-Paule Teulade-Fichou<sup>8</sup>, Benoit Ballester <sup>2</sup> & Marcel Méchali<sup>1</sup>

Genome-wide studies of DNA replication origins revealed that origins preferentially associate with an Origin G-rich Repeated Element (OGRE), potentially forming G-quadruplexes (G4). Here, we functionally address their requirements for DNA replication initiation in a series of independent approaches. Deletion of the OGRE/G4 sequence strongly decreased the corresponding origin activity. Conversely, the insertion of an OGRE/G4 element created a new replication origin. This element also promoted replication of episomal EBV vectors lacking the viral origin, but not if the OGRE/G4 sequence was deleted. A potent G4 ligand, PhenDC3, stabilized G4s but did not alter the global origin activity. However, a set of new, G4-associated origins was created, whereas suppressed origins were largely G4-free. In vitro *Xenopus laevis* replication systems showed that OGRE/G4 sequences are involved in the activation of DNA replication, but not in the pre-replication complex formation. Altogether, these results converge to the functional importance of OGRE/G4 elements in DNA replication initiation.

<sup>1</sup>Institute of Human Genetics, CNRS-University of Montpellier, 141 rue de la Cardonille, 34396 Montpellier, France. <sup>2</sup>Aix Marseille Univ, INSERM, TAGC, Marseille, France. <sup>3</sup>Balasubramanian group, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. <sup>4</sup>ARNA Laboratory, Université de Bordeaux, Inserm U1212, CNRS UMR5320, Institut Européen de Chimie Biologie (IECB), Pessac 33607, France. <sup>5</sup>Institut Curie, CNRS UMR9187, Inserm U1196, Université Paris Saclay, Orsay, France. <sup>6</sup>Research Unit Gene Vectors, Helmholtz Zentrum München (GmbH), German Research Center for Environmental Health, Marchioninistraße 25, 81377 Munich, Germany. <sup>7</sup>Monoclonal Antibody Core Facility & Research Group, Institute for Diabetes and Obesity, Helmholtz Zentrum München, Ingolstädter Landstrasse, 85764 Neuherberg, Germany. <sup>8</sup>CMIB Laboratory, Institut Curie, Campus Universitaire, Bat 110, 91405 Orsay, France. <sup>9</sup>Present address: Centre de Recherche en Cancérologie de Marseille 27 Boulevard Lei Roure, 13273 Marseille, France. <sup>10</sup>These authors contributed equally: Marie Artufel, Antoine Aze, Philippe Coulombe. Correspondence and requests for materials should be addressed to B.B. (email: [benoit.ballester@inserm.fr](mailto:benoit.ballester@inserm.fr)) or to M.M. (email: [marcel.mechali@igh.cnrs.fr](mailto:marcel.mechali@igh.cnrs.fr))

In mammals, around 100,000 potential DNA replication origins (origins throughout the text) are distributed along chromosomes. However, only about 30% is activated in a cell, in an apparent stochastic way. This flexibility in origin choice is considered an important feature for the robustness of DNA replication, and for the adaptation to DNA replication stress and cell fates (for a review)<sup>1</sup>. The second main feature of metazoan origins is their sequence plasticity. Indeed, differently from *Saccharomyces cerevisiae* origins, metazoan origins do not have a unique conserved consensus element. Some genetic and epigenetic characteristics have been identified in the vicinity of origins, but none can be considered to be a universal feature of metazoan origins. Among these features, the Origin G-rich Repeated Element (OGRE) is present in more than 60% of origins, in fly, mouse, and human cells<sup>2–6</sup>. This element can potentially form a G quadruplex (G4) structure (thereafter, such sequence elements are defined as OGRE/G4), and it is upstream of the initiation site (IS) of DNA synthesis, at an average distance of 250–300 bp. This localization could be compatible with the position of the pre-replication complex (pre-RC), and is associated with a nucleosome-free region<sup>4</sup>. The presence of similar elements at human origins has been detected using a different method than those used for mouse<sup>2,6</sup> and chicken cells<sup>7</sup>, and it was shown that proteins involved in DNA synthesis initiation, ORC<sup>8</sup>, MTBP<sup>9</sup> and MCM2–7<sup>10</sup> are also associated to such elements. A functional evidence for the use of this element was reported in chicken cells in a 1.1 kb fragment of the  $\beta$ -globin replication origin flanked by an HS4 insulator included close to a blasticidin resistance transgene under the control of the strong actin promoter<sup>7</sup>. However, it is unclear whether this result can be translated to other model systems, and no analysis has been done so far on a natural replication origin, at its original site or at an ectopic position.

Here we used various experimental approaches to determine whether OGRE/G4 is a functional element at metazoan origins. First, using an in vivo genetic approach at an endogenous locus, we showed that deletion of this motif strongly reduced origin activity in mouse cells. Moreover, an OGRE/G4-containing sequence introduced in an ectopic origin-free region promoted the establishment of a new functional origin. Second, we showed that a plasmid containing an origin with an OGRE/G4 element can replicate in HEK293 cells that express EBNA1 almost as efficiently as plasmids containing the Epstein-Barr virus (EBV) origin OriP, and that deletion of the OGRE/G4 element strongly reduces its replication efficiency. Third, we analyzed the influence of PhenDC3, a known G4 ligand, on origin firing efficiency genome-wide. Fourth, we performed competition experiments in in vitro systems of DNA replication derived from *Xenopus laevis* eggs, and found that G4-forming sequences are competitors that strongly affect DNA replication initiation.

Altogether, all our results converge to the conclusion that G-rich elements, including the OGRE/G4 motif, are functionally important for origin activity.

## Results

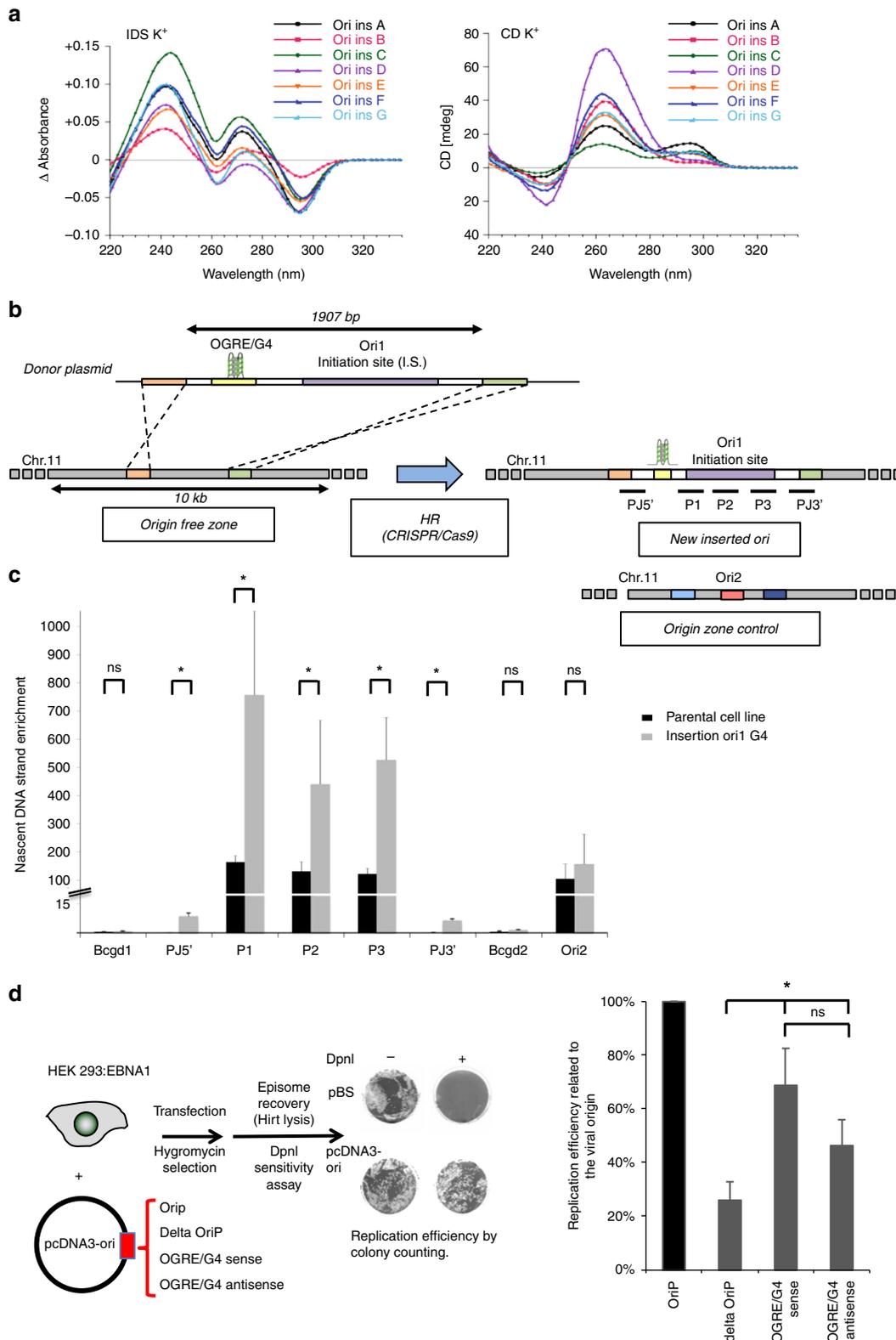
**OGRE/G4 elements can form G4 in vitro.** We first asked whether the OGRE/G4 motif could form G4 in vitro. Origins were identified from which cells by purification of Short RNA-primed Nascent Strands (SNS), a procedure that we and others repeatedly found to be accurate for origin analysis in *Drosophila melanogaster*<sup>5</sup>, mouse<sup>2–4</sup> *Arabidopsis thaliana*<sup>11</sup>, *Caenorhabditis elegans*<sup>12</sup>, chicken<sup>7</sup>, and human cells<sup>13–15</sup>, and the results of which were confirmed by different approaches<sup>6,10,14,15</sup>. Supplementary Fig. 1 summarizes this procedure (detailed in “Methods” section), and shows the controls used for this analysis.

We tested the capacity of G4 formation by sequences found in the origin vicinity using isothermal difference spectra (IDS) and circular dichroism (CD). To test their propensity to form a G4-structure, we selected origins in different chromatin domains, transcription status and replication activity. Because each sequence needed to be individually synthesized and tested by CD and IDS, we did a selection of 7 origins. The bioinformatics prediction for a potential of G4-structure was first tested at the bioinformatical level, using the G4H algorithm (similar results were obtained with the Quadparser software), and indicated a high capacity for G4-formation for all tested sequenced (Fig. 1a and Supplementary Table 1). Circular dichroism (CD) is a highly sensitive assay, which can determine the conformational state of quadruplex structures<sup>16</sup>. Isothermal differential spectra (IDS) are obtained using a method derived from that for thermal denaturation spectra<sup>17</sup>; they provide information on the nature of the folded structure. Both assays showed that all these sequences exhibited hallmarks of quadruplex formation, as shown by the strong negative peak around 295 nm and the two positive peaks around 240 and 273 nm for IDS (Fig. 1a, left panel), and the strong positive peak around 260 nm with CD (Fig. 1a, right panel). Such data suggested a predominantly parallel quadruplex conformation for all sequences and confirmed G4 formation by these sequences.

**OGRE/G4 elements confer replication origin activity.** We then selected a strong and reproducible origin that was present in all our five independent experimental replicates (Ori 1, Supplementary Table 1; Supplementary Fig. 2A shows the raw data in our replicates). The replication origin positions were defined in a genome-wide manner using MACS2 and SICER peaks calling softwares, as previously described<sup>4</sup>. The origin initiation site is the highest NS-enrichment score over the initiation region. The OGRE/G4 motif was located 240 nt upstream of the IS (Fig. 1b and Supplementary Fig. 2A), in agreement with previous results in mouse cells<sup>4</sup>. After insertion of a 1907 bp fragment that included the OGRE/G4-containing Ori1 into a large region devoid of replication or transcription activity (Fig. 1b, Supplementary Fig. 2A–C and “Methods” section), we tested replication activity by SNS purification followed by qPCR with primers for the inserted origin sequence (Supplementary Table 2 and “Methods” section). The replication profiles showed that Ori1 was active at the ectopic position (Fig. 1c). As the inserted sequence was identical to the original sequence, the origin activity observed after the insertion was around twice the activity measured in parental cells. Conversely, the activity of another origin on chromosome 11 (external origin, Ori2) did not change (Fig. 1c).

To functionally assess the importance of the OGRE/G4 motif, we also used another experimental system based on the replication of episomal DNA in mammalian cells. This episomal plasmid harbors the EBV origin OriP that is recognized by the viral protein EBNA1<sup>18</sup>. OriP is a bipartite element consisting of the family of repeats (FR) and the dyad symmetry (DS) element. Both are recognized by EBNA1, favoring the mitotic segregation of the episome and DNA replication during S phase respectively<sup>19</sup>. Interestingly, replication occurs ORC dependently once per cell cycle in synchrony with chromosome replication<sup>20–22</sup>.

After transient transfection of different episomal plasmids (Fig. 1d, left panel) in HEK293 cells that stably express EBNA1, we analyzed episomal DNA replication by DpnI digestion/transformation (Fig. 1d and “Methods” section). DS deletion ( $\Delta$ OriP) strongly inhibited episomal DNA replication, showing the requirement of an active origin in this system (Fig. 1d,



right panel). Insertion of a 500 bp mouse OGRE/G4-containing origin (Ori2; Supplementary Table 1) at the place of OriP (Fig. 1d) also promoted episomal DNA replication almost as efficiently as the viral origin. As previously shown, the OGRE/G4 presence is orientated relative to the initiation site, as initiation occurs always downstream to the OGRE/G4<sup>3,4</sup>. So, when the antisense sequence is used, the initiation site will be in the other

direction. In this orientation the origin is still active, as expected, although slightly less possibly because of a different chromatin environment in the reverse direction.

From these results, obtained in two different in vivo systems and using different methods to analyze origin activity, we concluded that an OGRE/G4-containing origin can function ectopically in the genome and also in episomal plasmids.

**Fig. 1** Creation of an ectopic DNA replication origin. **a** Isothermal differential spectra (IDS; left panel) and circular dichroism spectra (CD; right panel) of potential OGRE/G4 sequences found in the vicinity of replication origins. All tested sequences form G4 structures, as indicated by the strong negative peak around 295 nm and the two positive peaks around 240 and 273 nm (for IDS), and the strong positive peak around 260 nm (CD). The sequences are provided in Supplementary Table 1. **b** Ori1 that contains an OGRE/G4 element 240 bp upstream of the DNA replication initiation site (IS) was inserted by Cas9-stimulated homologous recombination into an origin-free region on chromosome 11 in NIH 3T3 mouse cells. The insertion of the 1907 bp fragment (marked in violet) occurred thanks to the two 500 bp homology arms (orange and green) present on the insertion template. The position of the primers (P) (sequences in Supplementary Table 2) used for the analysis of origin activity is also shown. **c** Ori1 activity in parental (control in black) and recombinant NIH 3T3 cells (in grey). As expected, a two-fold increase in DNA replication activity was detected in recombinant cells compared with parental cells, whereas the external origin Ori2 exhibited the same replication activity in both cell lines. Note that SNS activity was also detected at the 5' and 3' junctions of the insertion site, but not in the corresponding control regions. The background control regions Bcgd1 and 2 are located in origin-free regions; results are the mean  $\pm$  SD of 3 independent experiments;  $p$  values were obtained using the two-tailed Student's  $t$  test;  $*p \leq 0.05$ ,  $p > 0.05$ . **d** Analysis of DNA replication, using the DpnI digestion method and colony counting ("Methods" section), in an EBV episomal plasmid transfected in HEK293 cells that express EBNA1. DNA replication activity was assayed using EBV episomal plasmids that carry or not (Delta) the OriP origin, or a 500 bp fragment of Ori2 containing an OGRE/G4 element in the sense or antisense orientation. Results are the mean  $\pm$  SD of 3–7 independent experiments;  $p$  values were obtained using the two-tailed Student's  $t$  test;  $*p \leq 0.05$ ; ns not significant,  $p > 0.05$

### Deletion of the OGRE/G4 inhibits replication origin activity.

To further confirm that the potential formation of a G4 is important for the origin functionality, we deleted the endogenous OGRE/G4 sequence in Ori1. Co-expression of the Cas9 nickase and two gRNAs targeting this sequence led to the formation of a double-strand break and the subsequent deletion of the targeted sequence (Fig. 2a, "Methods" section and Supplementary Fig. 3A). The strong peak (G4H score  $> 2$ ) observed in the wild type sequence with G4-Hunter (a tool to predict the propensity of a sequence to form G4) disappeared for both mutated alleles (no signal above 1), strongly suggesting that our targeted deletion removed the putative G4-forming sequence at this locus (Fig. 2b). In order to confirm these predictions, we analysed the circular dichroism (CD) (Supplementary Fig. 3B, left panel) and isothermal differential spectra (IDS) (Supplementary Fig. 3B, right panel) of Ori1 wt sequence and 2 mutated alleles of Ori1. The results indicated a strong capacity of G4-formation by the wt sequence with a strong positive peak around 260 nm on CD spectrum, and a strong negative peak around 295 nm and the two typical positive peaks around 240 and 273 nm on IDS spectrum. In agreement with the bioinformatics predictions these hallmarks of G4-formation are lost in mutated Ori1 alleles. It is noteworthy that the bioinformatics predictions gave a very accurate prediction of G4-forming potential that was confirmed by in vitro CD and IDS spectra analysis for all tested sequences (Fig. 1a, Supplementary Fig. 3B, C).

Quantification of the origin activity by SNS purification and qPCR analysis showed that in mutant cells, Ori1 replication activity was decreased by 85%, but not that of an external origin (Ori2), also located on chromosome 11 (Fig. 2c). The transcription levels of the *Rai1* gene, associated with Ori1, and of the *Actb* (actin) and *Gapdh* controls were only slightly affected (Fig. 2d), making unlikely an indirect effect due to a transcriptional activity change.

Similarly, deletion of the OGRE/G4 sequence in the episomal vector strongly inhibited episomal DNA replication (Fig. 2e). Randomization of the OGRE/G4 sequence also decreased origin efficiency, suggesting that, at least for Ori2, G-richness *per se* is not sufficient and that G4 formation is an important feature (Fig. 2e). Additionally, to confirm the capacity of G4-formation by Ori2 and its absence in Randomised Ori2 we analysed the circular dichroism (CD) (Supplementary Fig. 3C, left panel) and isothermal differential spectra (IDS) (Supplementary Fig. 3C, right panel) of Ori2 wt sequence and Randomised Ori2. The results unambiguously showed a strong G4-forming potential in the wt sequence that was completely lost in Random mutant.

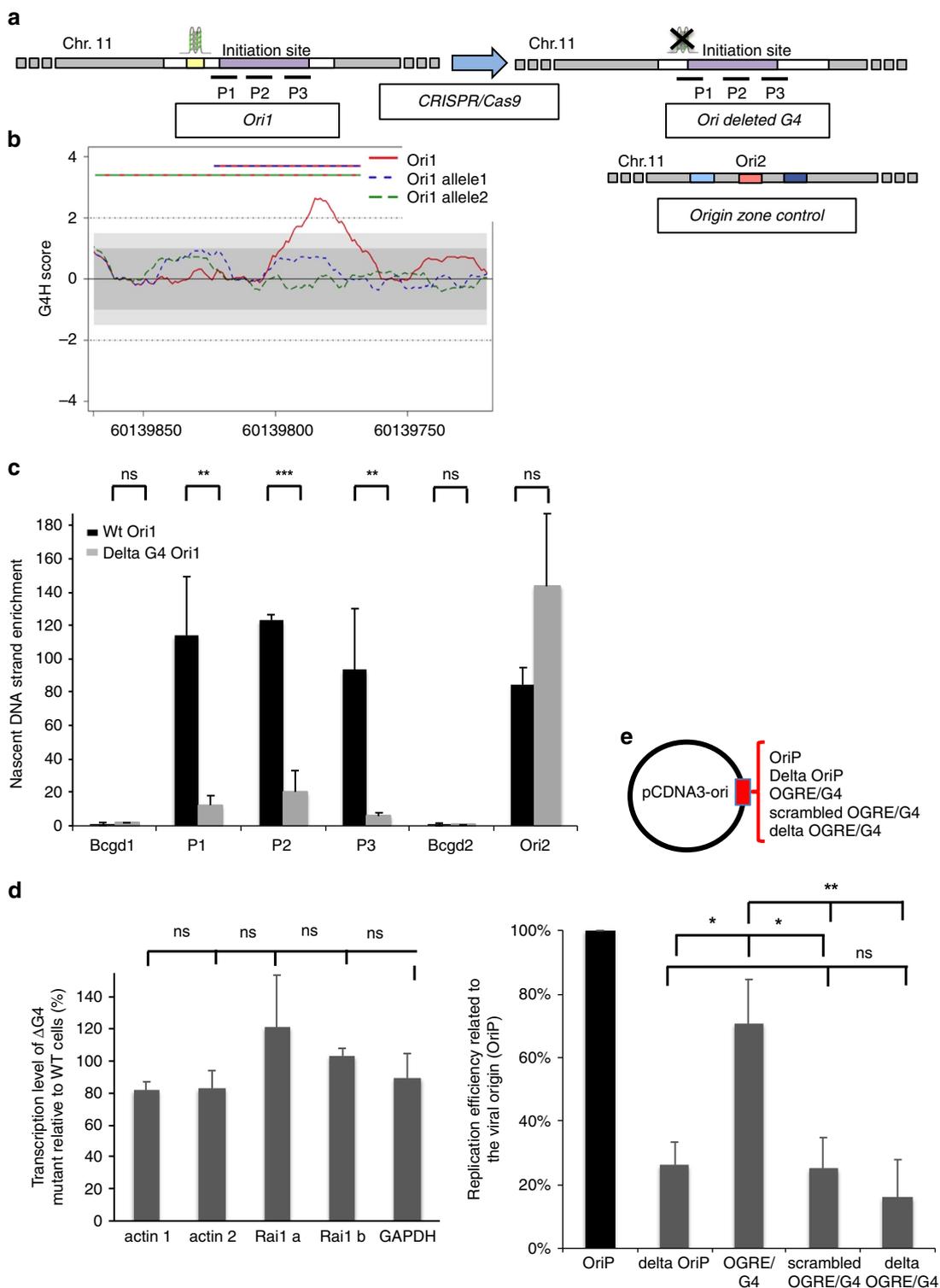
Altogether, these functional studies indicate that the OGRE/G4 element located upstream of Ori1 is functionally active and positively contributes to origin activity.

**G4-stabilization increases G4-associated origins firing.** To better understand the importance of OGRE/G4 elements, we investigated genome-wide whether G4 stabilization could affect origin activity in mouse embryonic stem (ES) cells. We used PhenDC3 (Fig. 3a), a bisquinolinium compound that has high affinity for G4 and that shows an exceptional selectivity for G-quadruplexes<sup>23,24</sup> compared with duplexes, as indicated by the increase in melting temperature ( $\Delta T_{1/2}$ ; stabilization) of seven different quadruplexes, but not for the control duplex (FdxT) (Supplementary Fig. 4A, and Supplementary Table 3).

Compared with control ES cells, incubation with 10  $\mu$ M PhenDC3 for 48 h, as previously described<sup>25</sup>, did not affect the cell cycle profile (Supplementary Fig. 4B), and the expression and phosphorylation of CHK1, a kinase involved in cell cycle progression and in the DNA damage checkpoint (Supplementary Fig. 4C). Conversely, CHK1 phosphorylation was induced by the genotoxic agents camptothecin (Cpt) and etoposide (Eto). The expression of OCT4, a pluripotency marker, also was not modified by PhenDC3 (Supplementary Fig. 4C).

We used a Volcano plot to identify statistically significant changes in replication origin activity (Fig. 3b). A Volcano plot visualizes the biological effect on the  $x$ -axis (Log2(fold change, FC)) and the statistical significance on the  $y$ -axis ( $-\log_{10}$ (false discovery rate, FDR)). This analysis allowed to define five origin classes according to their activity in response to PhenDC3: *insensitive*, *new*, *reinforced*, *reduced* or *suppressed* (Fig. 3b–e and Supplementary Table 4). Examples of origins belonging to these classes are shown in Fig. 3c, while the fold change in origin activity for each class is depicted in Supplementary Fig. 4D. Overall, we did not observe a substantial increase of origins in the presence of PhenDC3 (Supplementary Table 4). The heatmap (Fig. 3d) showing read density in the vicinity ( $\pm 7$  kb) of origins indicated that *reduced* and *suppressed* origins were situated in an origin-dense environment as opposed to *reinforced* and *new* origins. Origins that remained at the same position and with a similar activity (PhenDC3 *insensitive*) represented 77.9% of all origins (Fig. 3e). One possible hypothesis could be that formation of a G4 is not essential for the activity of most origins, but this is in contradiction with our functional analyses showing the requirement of the OGRE/G4 element for origin activity (Figs. 1 and 2). A second possibility is that most G4 were normally formed during origin assembly or activation with no need of further stabilization by PhenDC3. It was nevertheless plausible that the genetic, chromatin and transcriptional landscape also influence the activity of G4 origins (see later).

Incubation with PhenDC3 also led to a set of *new* origins (15.7% of all origins) with a level of activity comparable to that of *insensitive* origins (Fig. 3e). Two smaller origin classes were



represented by *reinforced* and *reduced* origins (0.6 and 0.7%, respectively). Origins *reduced* by PhenDC3 were initially particularly strong, among the top 10% of strongest origins (Fig. 3e). Overall, we observed that the ligand-mediated G4 stabilization led to a more uniform activity of all origins (Fig. 3e).

We used the RSAT peak-motifs program (see “Methods” section) to find a specific motif in these origin classes. This led to the de novo identification of a G-rich motif upstream the IS, similar to the OGRE/G4 element<sup>2,4</sup>, in all origin classes but for

the *suppressed* class (Fig. 4a and Supplementary Fig. 5A). *Suppressed* origins were G4-poor, and preferentially contained a GC-rich motif (Supplementary Fig. 5A), possibility reflecting the enrichment of these origins in GCI promoters and a link with transcription (see below). In *reduced* origins, the OGRE/G4 element was at almost 400 bp upstream of the IS (Fig. 4b). This could be a consequence of their localization close to a promoter. We concluded that G4 stabilization by PhenDC3 did not reveal any new motif in the replication origin repertoire, but led to the

**Fig. 2** OGRE/G4 deletion strongly decreases the DNA replication activity of an endogenous origin. **a** The OGRE/G4 sequence of an endogenous origin (Ori1) was deleted and the deletion was confirmed using a restriction site close to the targeted sequence (see “Methods” section). **b** G4 formation propensity profiling of the Ori1 sequence targeted for deletion. The Ori1 sequence is located on chromosome 11 and presents a strong peak in the G4-Hunter score profile (red line). Such peak is not present upon OGRE/G4 deletion (alleles 1 and 2, blue and green dotted lines, respectively), and no point above 1 or below  $-1$  is observed. This argues against the probability of G4 formation at this mutated locus. The striped lines on the top indicate the extent of the deletion in allele 1 (red and blue) and allele 2 (green and blue). **c** Nascent strand enrichment of Ori1 in parental NIH 3T3 cell line (black) and in mutant clones with the deletion (grey). Replication activity was strongly decreased after deletion of the OGRE/G4 sequence, whereas the activity of the external origin (Ori 2) did not vary. The background control regions Bcgd1 and 2 are located in origin-free regions. Results are the mean  $\pm$  SD of 3 independent experiments;  $p$  values were obtained using the two-tailed Student’s  $t$  test; \* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ ,  $p > 0.05$ . **d** Deletion of the OGRE/G4 did not affect the transcription level of the Rai1 gene, associated with Ori1. As a control, the housekeeping genes Actb and Gapdh were used. Results are the mean  $\pm$  SD of 3 independent experiments;  $p$  values were obtained using the two-tailed Student’s  $t$  test;  $p > 0.05$ . Primer sequences are in Supplementary Table 2. **e** DNA replication activity was assessed as in Fig. 1c with the EBV origin, or with the 500 bp OGRE/G4 element of Ori2, or after scrambling or deletion of the same OGRE/G4 sequence. Results are the mean  $\pm$  SD of 4–5 independent experiments;  $p$  values were obtained using the two-tailed Student’s  $t$  test; \* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ;  $p > 0.05$ . Note that data presented for episome containing delta oriP and OGRE/G4-containing origin were performed independently from results presented in the Fig. 1; ns not significant:  $p > 0.05$

suppression of a discrete origin population that lack the OGRE/G4 sequence. These origins were in origin-dense regions, and their suppression might compensate the appearance of new OGRE/G4-containing origins, favored by their PhenDC3-mediated stabilization.

We experimentally tested G4 formation in vitro in a subset of *new* origins using CD and IDS, as previously described (Fig. 1a and Supplementary Table 1 for the full list of tested sequences). All these sequences exhibited the hallmarks of quadruplex formation (Fig. 4c). The presence of a minor peak around 295 nm may indicated the formation of alternative folds (possibly anti-parallel G4 structures) for some sequences.

We then asked whether the five origin classes correlated with putative G4 predicted by the G4-Hunter (G4H)<sup>26</sup> and Quadparser (QP) algorithms<sup>27</sup> (Fig. 4d and Supplementary Table 4). The bioinformatics analysis gave highly accurate predictions of G4-forming potential that was confirmed by CD and IDS analysis for several wt and mutant sequences (Fig. 1a, Supplementary Fig. 3B, C). Using stringent parameters, 490,971 G4 were predicted by Quadparser (G-track size min = 3; parameters loop size min = 1, max = 7, Gs permitted in the loop), and 568,806 by G4-Hunter (threshold = 2, window size = 25). Analysis of G4 distribution in each origin class gave similar results with both software programs. This analysis showed no difference in G4 score distribution among classes (Supplementary Fig. 5B; for simplicity, only the results with G4-Hunter are shown), which indicates that G4 strength does not explain our observations.

Moreover, we did not find any significant correlation between the length of the OGRE/G4 sequence and the different origin classes (Supplementary Fig. 5A, C), but we detected a slight global effect of the number of OGRE/G4 motifs present close to the IS (Supplementary Fig. 5D).

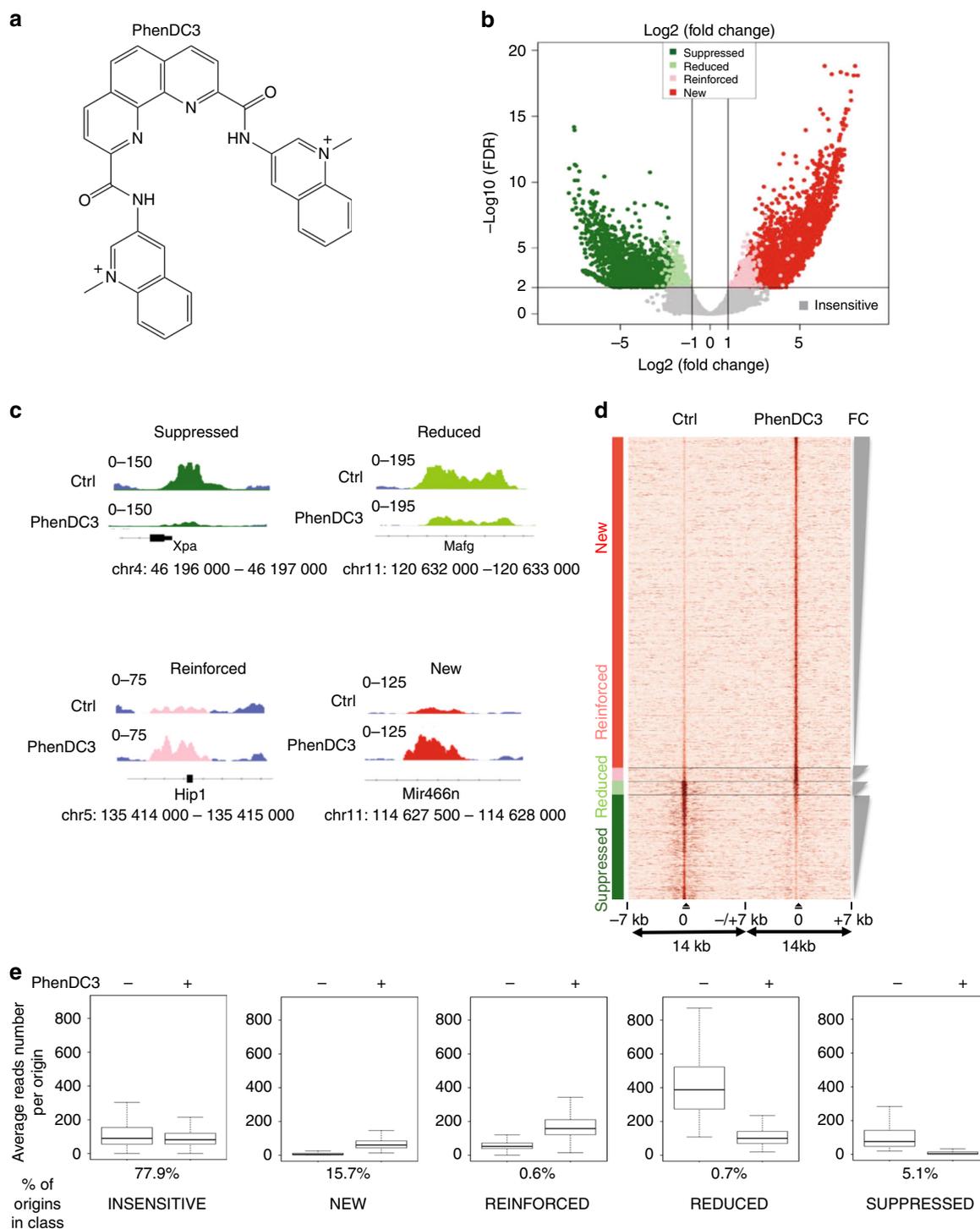
Finally, to confirm the functional link between PhenDC3 effect and the OGRE/G4 motifs, we used an indirect FRET melting competition assays with OGRE/G4 oligonucleotide sequences from the *insensitive* class (which were the same as tested for G4-formation using CD et IDS) and *new* origin classes as well as unlabeled positive (G4) and negative controls (single- or double-strand oligonucleotides) (Supplementary Table 1). These sequences were added to a mixture containing a double-fluorescently labeled G4 forming sequence (F21T) corresponding to the human telomeric motif, in the presence or absence of PhenDC3. PhenDC3 bound to F21T and increased its melting temperature in a concentration-dependent manner ( $\Delta T_m = +29^\circ\text{C}$  at  $1\ \mu\text{M}$  and  $\approx +18^\circ\text{C}$  at  $0.5\ \mu\text{M}$ ; Fig. 4e) when no competitor was present. Negative control competitors, unable to bind to PhenDC3 (dT30 and DS26; single- and double-strands, respectively) did not affect this stabilization, as expected given the

high specificity of PhenDC3 for G4 structures. Conversely, the strong decrease in stabilization observed after addition of origin sequences confirmed the recognition by PhenDC3 of the OGRE/G4 motifs in these origins that acted as strong competitors for PhenDC3 (Fig. 4e). We concluded that PhenDC3 displays high affinity for both *insensitive* and *new* origins, confirming the functional link between PhenDC3 incubation and the observed changes in replication activity.

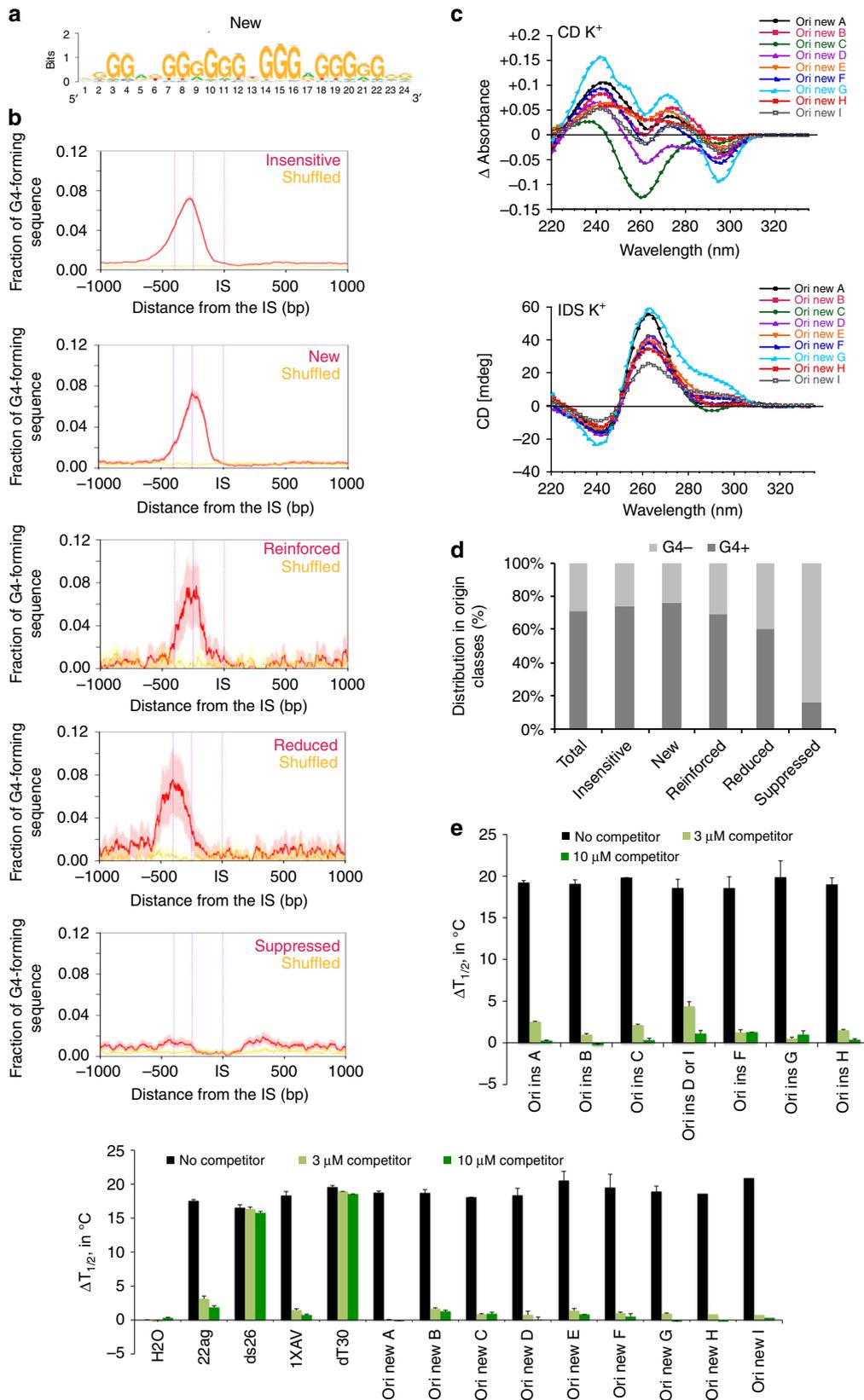
#### Transcription and not G4 govern replication at promoters.

Analysis of the genomic location showed that overall, origins were enriched at gene regions, compared with intergenic regions (Fig. 5a), as previously widely observed<sup>28</sup> and references herein). Remarkably, *suppressed* and *reduced* origins were highly enriched at promoter regions, whereas the other origin classes were mostly absent from promoters and evenly distributed between transcribed and intragenic regions (Fig. 5b, random origins as dotted lines, and control in Supplementary Fig. 6A), confirming our previous results. Next, we asked whether the five origin classes defined in this study were associated with specific chromatin signatures. Pearson correlation analysis using BEDTools<sup>29</sup> (see “Methods” section) revealed that *suppressed* and *reduced* origins were strongly correlated with chromatin marks associated with active transcription and with bivalent epigenetic marks (Fig. 5c, control randomized regions in Supplementary Fig. 6B, and reference data in Supplementary Table 5). *Suppressed* and *reduced* origins were also associated with several transcription factors, further confirming the promoter location of these origins. This result also explains why origins from the reduced class exhibited stronger replication activity before G4-stabilisation. These origins were quite strong because of the presence of both G4 and active transcription. The decreased transcription activity at these origins upon G4 stabilisation decreased the stimulating effect of transcription of these origins. *Reinforced* origins were only slightly correlated with enhancer marks (Fig. 5c). Moreover, we observed a strong link between the formation of *new* origins and regions poor in epigenetic marks, but enriched in G4-forming fully methylated sequences (Fig. 5c).

To further interpret these results, we analyzed the transcriptional output associated with each origin class by RNA-seq analysis of control and PhenDC3-treated ES cell samples (“Methods” section) followed by identification of genes that were differentially expressed in each class using the DESeq2 algorithm (see “Methods” section). We computed the enrichment set using the genes associated with each origin class and by considering the origin localization at the promoter (TSS  $\pm$  2 kb, left panel) or within the transcribed regions (TSS + 10 kb). We found that origin activity tended to follow the



**Fig. 3** Changes in the replication origin repertoire upon G4 stabilization by PhenDC3. **a** PhenDC3 formula. **b** Volcano plot of origins affected by incubation of mouse ES cells with PhenDC3. After identification of the bound sites in all SNS-seq samples, differential binding analysis was performed. For each origin, the corrected *p* values (false discovery rates,  $-\log_{10}(\text{FDR})$ ) and the  $\log_2$  fold change (FC) of control and PhenDC3-treated samples were plotted. The horizontal and vertical lines correspond to the thresholds for detecting differential origins. On the basis of the FC and peak reproducibility, origins were classified in five different classes, according to PhenDC3 effect (*suppressed*, *reduced*, *reinforced*, *new*, and *insensitive*), as described in “Methods” section. **c** Examples of the activity of origins in the indicated classes after incubation with PhenDC3 or in control cells. The corresponding genomic region is indicated and the origin color is according to the corresponding class in the Volcano plot. **d** Heatmap showing the read densities in origins affected by G4 stabilization (PhenDC3-treated vs. Control). The heatmap indicates the signal strength (number of reads) and density around each origin and was performed on 7 kb regions on each side of origins, as previously described<sup>4</sup>. The intensity (brown) is proportional to the read counts per 100 bp bins. Origins were sorted on the basis of the FC in signal strength. **e** Activity of origins (reads number) in each class in control (–) and PhenDC3-treated (+) cells



transcriptional output. For instance, *suppressed* origins found at promoters were significantly associated with gene repression (Fig. 5d), as well as *reduced* origins. Conversely, *new* and *reinforced* origins found at promoters tended to be associated

with upregulated genes. This is also in agreement with the observation that origins close to TSS are usually highly active, and are downregulated when transcription decreases<sup>1,28,30</sup>. We did not detect any correlation between replication and

**Fig. 4** Nature of the OGRE/G4 in the different origin classes. **a** De novo motif found as the most representative in the new origin class using the RSAT suite<sup>60</sup>. For motifs found in the other classes see Supplementary Fig. 5A. **b** Fraction of OGRE/G4 sequences in function of the distance from the IS. The OGRE/G4 motif forms a relatively sharp peak upstream the IS at an average distance of 250 bp in all origin classes, but for the suppressed class. The CL95% is shown in pink. The fraction of OGRE/G4 sequences in shuffled regions and their CL 95% is shown in yellow and light yellow, respectively. **c** Isothermal differential spectra (IDS; upper panel) and circular dichroism spectra (CD; lower panel) of potential OGRE/G4 sequences associated with the new class of replication origins. All tested sequences form G4 structures, as indicated by the strong negative peak around 295 nm and the two positive peaks around 240 and 273 nm (for IDS, top panel), and the strong positive peak around 260 nm by CD (bottom panel). The IDS suggest that the tested sequences form predominantly G4 in parallel conformation. The possibility to adopt alternative folds, such as anti-parallel G4 structures, for some sequences is indicated by a minor peak around 295 nm. The sequences are provided in Supplementary Table 1. **d** Association of origins with OGRE/G4 motifs in the different classes. Insensitive, new, enforced and reduced origins are mainly G4-associated, but not suppressed origins. **e** FRET competition assays in which stabilization ( $\Delta T_{1/2}$ , in °C) of the human telomeric quadruplex F21T by 0.5  $\mu$ M PhenDC3 was analyzed in the absence (black bars), or in the presence of G-rich sequences from insensitive (upper panel) and new (lower panel) origins (3 or 10  $\mu$ M strand concentration; dark green and light green bars, respectively), of positive (22Ag, 1XAV, both forming G4 structures), and negative (ds26 and dT30 are double- and single-stranded controls, respectively) control sequences. The means were obtained in independent experiments  $\pm$  SD. Efficient competition by quadruplex-forming oligonucleotides is evidenced by a sharp drop in stabilization. The origin oligonucleotide sequences are provided in Supplementary Table 1

transcription changes for origins situated in transcribed regions (Fig. 5d, right panel).

We concluded that *i*) replication origins are enriched in transcribed regions, including promoter; and *ii*) origins situated at promoters are often devoid of OGRE/G4 sequences, and their firing activity strongly depends on the transcription level. Conversely, G4 stabilization might facilitate origin firing in non-genic regions that are less prone to chromatin opening, or spontaneous G4-formation, such as fully methylated regions. In these regions, OGRE/G4 might help replication origin activity through its two main features: the presence of single-stranded DNA in the strand opposite to the G4, and its ability to exclude nucleosomes, and to favor a less energetically demanding origin activity in transcriptionally silent regions.

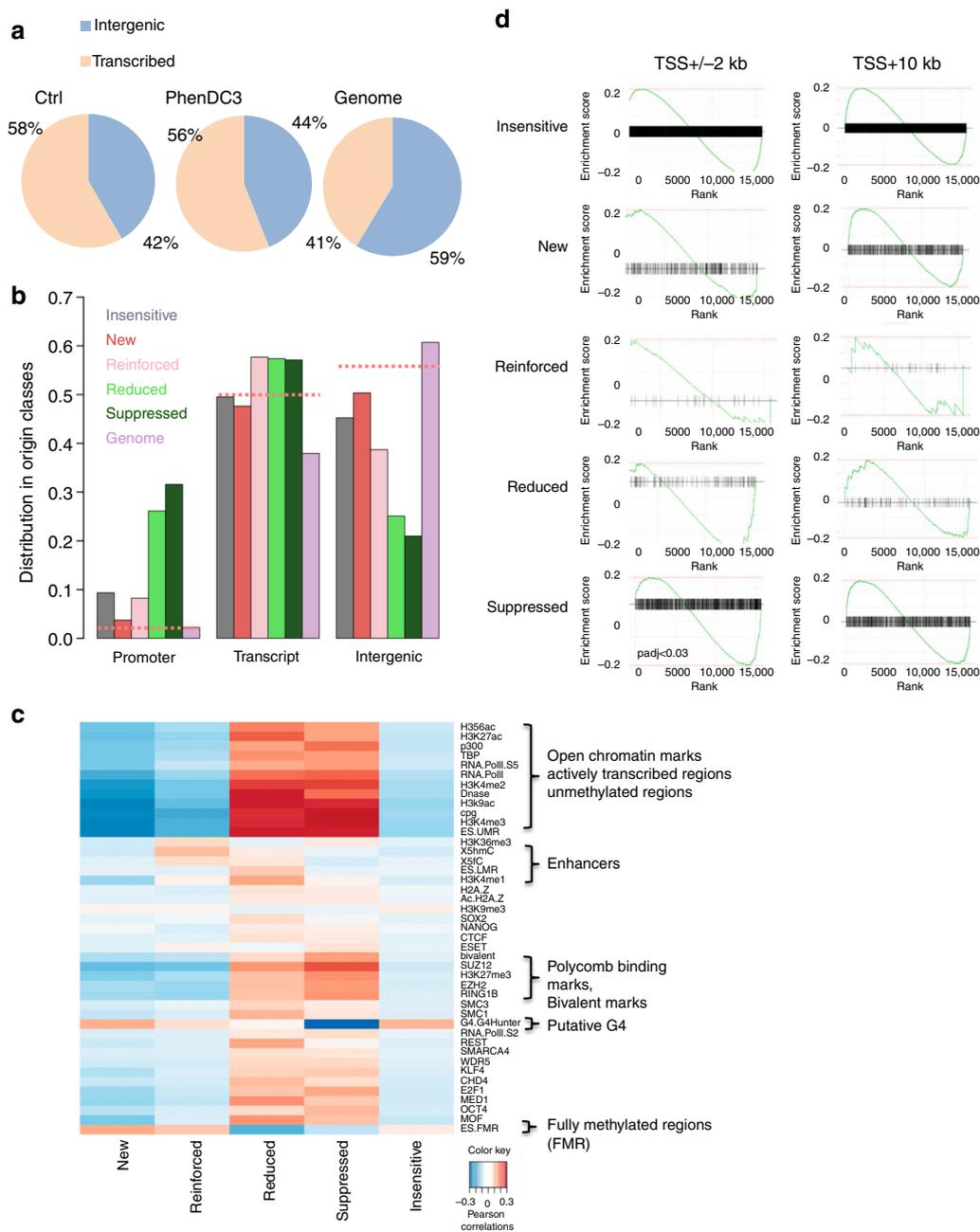
#### G4-forming oligonucleotides compete for replication factors.

Initiation of DNA replication is a two-step process. First (i.e., replication licensing), pre-RCs are assembled at origins and this includes the binding of ORC, CDC6, CDT1 and the MCM helicase. Then, the MCM helicase is activated and allows the recruitment of the DNA polymerase machinery. To determine whether OGRE/G4 elements could be potential binding sites for proteins involved in these steps, we performed classical oligonucleotide competition experiments in *Xenopus laevis* low-speed egg extracts (LSE). *Xenopus* LSE is a well defined cell-free system that faithfully reproduce DNA replication in vitro<sup>31</sup>. This reaction is entirely transcription-independent, thus excluding any influence by the transcription process on the assay. Oligonucleotides similar to the endogenous target DNA sequence should compete for the replication activity as opposed to oligonucleotides which are not related to the target sequence. To test whether OGRE/G4 oligonucleotide templates compete for factors involved in DNA synthesis on sperm nuclear chromatin (Fig. 6a), we incubated *X. laevis* LSEs with 80-mer oligonucleotides that contained an OGRE/G4 sequence (from Ori1 used in the CRISPR/Cas9 experiments), or a sequence with the same G content but randomized (random oligonucleotide), or an AT-rich sequence (Supplementary Table 6), or water (mock), or sonicated salmon sperm DNA. The kinetics of nuclear DNA replication (oligonucleotides do not replicate in the extract) were comparable in mock-treated extracts and after addition of sonicated salmon sperm DNA. DNA replication was slightly delayed by incubation with random and AT-rich oligonucleotides, whereas it was nearly abolished by OGRE/G4 oligonucleotides (Fig. 6a, and quantification in Fig. 6b). Differently from LSEs, *X. laevis* high-speed egg extracts (HSE), in which nuclear membranes have been removed, cannot initiate dsDNA replication. However, they can perform all the reactions occurring

during complementary DNA strand synthesis, as tested with ssM13 DNA as template<sup>32</sup> including RNA priming, elongation and ligation of Okazaki fragments, and chromatin assembly coupled to DNA synthesis. In these extracts, DNA synthesis was not affected by pre-incubation with OGRE/G4 or random oligonucleotides (Fig. 6c). We concluded that OGRE/G4 oligonucleotides compete specifically with replication initiation, and have little or no effect on the subsequent steps.

**G4 are involved in replication origin firing step.** It is unlikely that OGRE/G4 oligonucleotides inhibit DNA replication through checkpoint activation because the DNA damage checkpoint is deficient in *X. laevis* early embryos<sup>33,34</sup>. In agreement, OGRE/G4 oligonucleotides did not induce CHK1 phosphorylation in our in vitro conditions (Supplementary Fig. 7A), differently from incubation with pApT at a concentration that mimics post-midblastula transition conditions known to induce the checkpoint, while pCpG do not<sup>35</sup> (Supplementary Fig. 7A, lane 5). Moreover, caffeine, a checkpoint inhibitor, did not rescue the inhibition of DNA replication by OGRE/G4 oligonucleotides (Supplementary Fig. 7B), whereas it did in a control experiment where DNA replication was inhibited by aphidicolin (Supplementary Fig. 7C). Altogether, these findings show that checkpoint activation does not explain the inhibition of DNA replication by OGRE/G4 oligonucleotides.

We then investigated which replication initiation step was inhibited by exogenous G4 oligonucleotides. Pre-RC formation can be analyzed in *X. laevis* HSEs that allow this reaction, but not DNA synthesis initiation. Factors involved in origin recognition (ORC5), the recruitment of the MCM helicase onto DNA, (CDC6), and the MCM complex (MCM4) were similarly loaded on chromatin in mock-treated HSEs and in samples incubated with salmon sperm DNA, random oligonucleotides, or OGRE/G4 oligonucleotides (Fig. 6d). Formation of the nuclear membrane also was not affected, as shown by the chromatin recruitment of ELYS, a protein required for the formation of a functional nuclear membrane<sup>35,36</sup> (Fig. 6e). Conversely, the recruitment of CDC45, which is needed for DNA synthesis activation<sup>37</sup>, and of factors required for DNA synthesis initiation and for DNA strand elongation (RPA, and PCNA) was strongly decreased (Fig. 6e–f). These results suggest that OGRE/G4 oligonucleotides do not disturb the licensing step of DNA replication, but rather affect the conversion of the pre-RC into the DNA synthesis elongation complex. This result is in agreement with the recent finding that origin firing activity by Mdm2-binding protein (MTBP) in *X. laevis* and human cells is dependent on its G4-binding motif<sup>9</sup>.

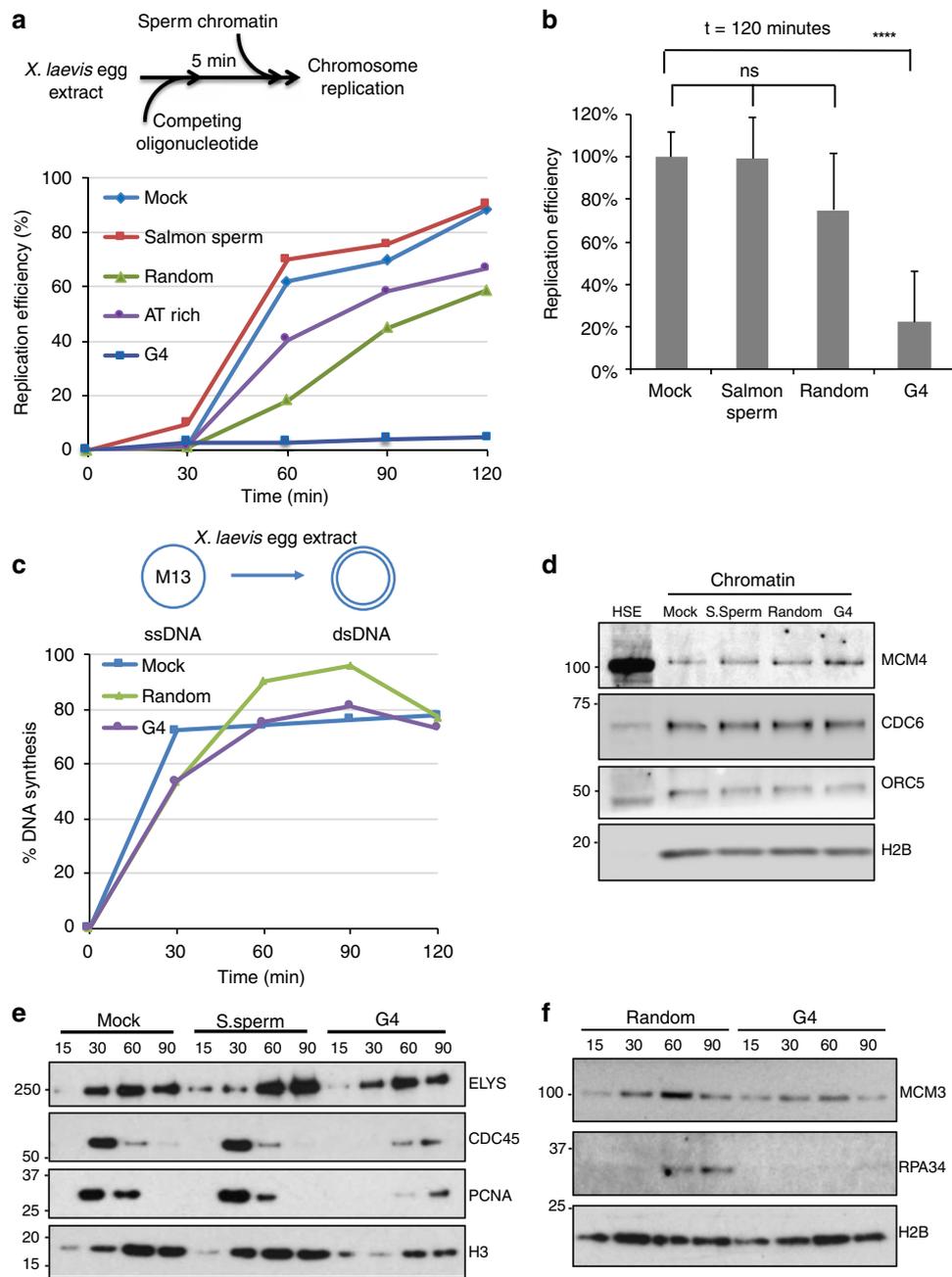


**Fig. 5** Transcription and epigenetic landscape in the different origin classes. **a** Venn diagrams showing the origin distribution between transcribed and intergenic regions in Control and PhenDC3-treated mES cells, and random distribution (Genome). Replication initiation sites are enriched in gene regions. **b** Genomic localization of the different classes of origins relative to transcription. Downregulated origins (suppressed and reduced) are mainly located at promoters. Random origins (dotted lines) are equally distributed in transcription-related regions. **c** Epigenetic marks associated with the different origin classes. All tested open chromatin marks were enriched around reduced and suppressed origins. New and reinforced origins were located mainly in highly methylated regions. **d** GSEA analysis of origins situated at promoters (TSS ± 2 kb, left panels) or in transcribed regions (TSS + 10 kb, right panels) for each class. A plot is drawn for each gene set. The x-axis of each plot represents differentially expressed genes ranked from upregulated (on the left) to downregulated (on the right). The enrichment score is indicated on the y-axis. The black horizontal bar indicates the genes present in the gene set. The highest enrichment score indicates the enrichment. If this score is on the left, the enrichment is higher for upregulated genes; if it is on the right, the enrichment is higher for the downregulated genes. Origins associated with the TSS follow the transcription changes upon G4 stabilization, whereas origins located in transcribed regions are insensitive to changes in transcription levels. At gene promoters, GSEA results show a significant association of downregulated genes with the suppressed origin class (after multi-testing correction using the Benjamini-Hochberg method, adjusted  $p < 0.03$ ). The other origin classes show a similar trend, without reaching significance

## Discussion

Genome-wide analyses of replication initiation profiles first highlighted that metazoan origins were enriched near CpG islands<sup>2,28,38,39</sup>. Then, the G-rich OGRE motif that could potentially form G4 was identified in the mouse and fly

genomes<sup>3,4</sup> and subsequently also in mouse<sup>4</sup>, chicken<sup>7</sup>, fly<sup>5</sup>, and human cells<sup>14,40</sup>. This element was discovered using the SNS purification system coupled with high-throughput sequencing (SNS-seq), which has currently the best resolution to map replication origins<sup>41</sup>. Moreover, G4 presence was detected also using  $\lambda$



**Fig. 6** At the activation step of DNA replication, OGRE/G4 elements compete for dsDNA but not for ssDNA replication. **a** Schematic representation of the replication kinetics of sperm nuclei in *X. laevis* low-speed egg extracts (LSE) in the presence of competing double-stranded (ds) oligonucleotides. LSEs were pre-incubated with competing oligonucleotides, sonicated salmon sperm DNA, or ultrapure H<sub>2</sub>O (mock) at 22 °C for 5 min. **b** Average DNA replication efficiency (mean + SD) of LSEs pre-incubated with competing oligonucleotides or controls ( $n = 6$  for mock/salmon sperm DNA pre-incubated extracts,  $n = 3$  for mock/random oligonucleotides). Total incubation time was 2 h.  $P$  values were obtained using the two-tailed Student's  $t$  test; \*\*\*\* $p < 0.0001$ ,  $p > 0.05$ . **c** Replication kinetics of ssM13 complementary DNA strand synthesis in *X. laevis* high-speed egg extracts (HSE) pre-incubated with the indicated competing oligonucleotides or H<sub>2</sub>O (mock). **d** Competition by OGRE/G4 oligonucleotides does not affect pre-RC formation. Sperm nuclei were added to HSEs incubated with H<sub>2</sub>O (mock), sonicated salmon sperm DNA (S.Sperm), or the indicated oligonucleotides. Chromatin was isolated and immunoblotted with the indicated antibodies. Histone H2B level was used as loading control. **e, f** Competition by OGRE/G4 oligonucleotides affects DNA replication activation. Time-course analysis of replication initiation factor recruitment to chromatin after addition of sperm nuclei to LSEs pre-incubated with H<sub>2</sub>O (mock), sonicated salmon sperm DNA (S. sperm), or competing oligonucleotides. At the indicated time points, chromatin was isolated and immunoblotted with the indicated antibodies; ns not significant

exonuclease-independent conditions<sup>6,15,40</sup>, and by genome-wide profiling of human replication origins after pulse labeling of SNS (Ini-Seq)<sup>6,14</sup>.

We used several complementary approaches to address the involvement of G-rich repeated elements and their potential to

form G4 structures in the activity of DNA replication origins. Our data confirm that such sequence elements are associated with the majority of active origins, and are localized just upstream of the initiation site. In vivo deletion or insertion of an OGRE/G4-containing wild type origin showed that the OGRE/G4 motif is

functionally active. This result was obtained using origins that are present in the mouse genome, as well as using recombinant episomal DNA.

In our ectopic assay, an OGRE/G4-containing fragment from an origin inserted in a region completely devoid of both DNA replication, transcription activity and G4-forming sequences led to the creation of a functional origin. Deletion of the OGRE/G4 element strongly decreased the activity of the origin. However, we cannot rule out that in other genomic regions, other features might stimulate or repress origin activity. Finally, we found that transcription activity of the gene associated with the origin remained unchanged upon origin deletion, indicating that the link between origin activity and transcription activity is not functionally compulsory. Moreover, our G4-stabilization assay suggests that this link is mostly limited to promoter regions.

OGRE/G4 elements exclude nucleosomes at mouse replication origins<sup>4</sup>. Nucleosome-free regions were also observed in *S. cerevisiae* origins<sup>42–45</sup>, although an AT-rich element characterizes their consensus origin-specific ARS element and also plays a role of nucleosome exclusion. OGRE/G4 elements might have a similar function in metazoans. Another possibility is that this sequence is the binding site for a replication initiation factor. In agreement with this hypothesis, recombinant ORC preferentially binds to G4-containing oligonucleotides<sup>8</sup>, as well as MTBP, partner of Treslin, that is involved in activation of origins of replication<sup>9</sup>. RIF1, a protein that regulates the timing of origin activation, also binds to putative G4-forming sequences<sup>46</sup>. Putative G4-forming sequences have also been observed at viral replication origins, such as the Kaposi sarcoma associated virus (KSHV) origin. This origin contains several G4 sequences and allows the stable maintenance of the viral episome in cells, and associates with ORC and MCM proteins<sup>47</sup>. Putative G4-forming sequences are also present at the EBV replication origin, to which EBNA1, the viral protein involved in origin recognition, binds<sup>48</sup>.

We used PhenDC3 as a G4-binding tool to reveal new G4-related features linked to replication origin activity. Incubation with PhenDC3 did not affect the activity of most origins, despite the presence of putative G4 sequences, suggesting that most origins do not need further stabilization by PhenDC3 for their activity. However, G4 stabilization increased the predisposition to become a replication origin for a subset of OGRE/G4-containing origins. These origins are mainly localized in non-coding regions that are poor in epigenetic marks and enriched in fully methylated regions. We propose that PhenDC3 might facilitate the formation of G4 structures in fully methylated regions that are less favorable to their formation<sup>49,50</sup>. The influence of DNA methylation status on G4-folding capacities has been very recently provided<sup>51</sup> using a G4-recognizing antibody which detected folded G4 structures in hypomethylated regions that overlap with DNMT1 binding sites. DNMT1 is a DNA methyltransferase that restores the DNA methylation pattern just after DNA replication. It has affinity for G4 structures, but surprisingly these structures inhibit its catalytic activity. In this way, DNMT1 can be concentrated in the vicinity of replication start sites and could immediately act on newly synthesized DNA after origin activation. Alternatively, PhenDC3 might facilitate the formation of G4 in heterochromatin structures, and therefore facilitate nucleosome exclusion and the formation of replication initiation complexes.

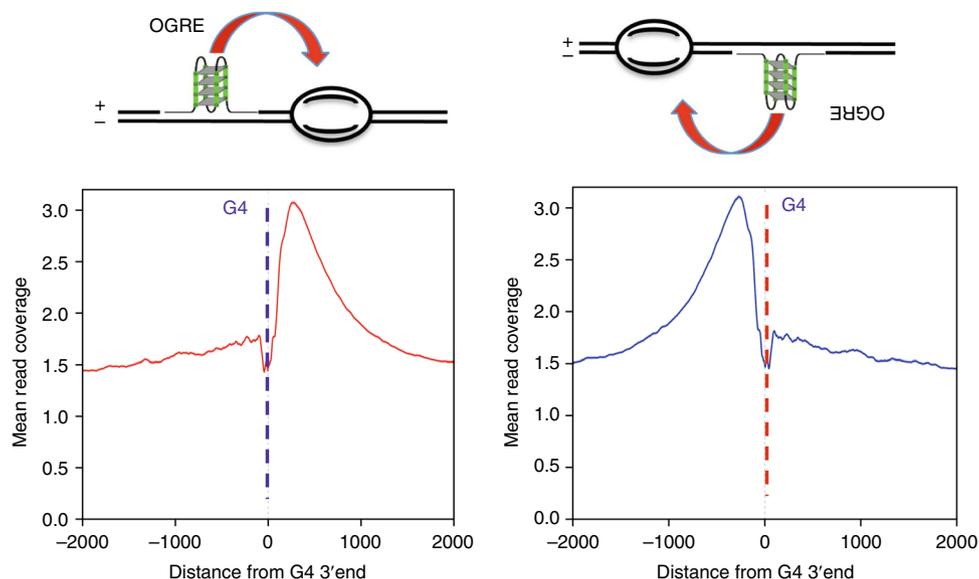
PhenDC3 incubation also led to the suppression of some origins that lack the OGRE/G4 element. These origins are found in promoters and are significantly associated with gene repression. We suggest that the replication activity of *suppressed* origins is mainly guided by transcription, and is not OGRE/G4-dependent. Transcription increases the activity of all origins close to a TSS when gene transcription is upregulated and decreases their

activity when transcription is downregulated. DNA replication can benefit from the open chromatin structure at gene promoters. However, the influence of transcription activity was limited to promoters, and transcription changes did not affect the activity of origins localized in gene bodies. This is in agreement with the observation that transcriptional silencing of the X chromosome does not induce changes in the strength or localization of the tested origins situated in gene bodies<sup>52</sup>. Finally, PhenDC3 incubation reduced the efficiency of a small fraction of origins (0.7%). These few origins were among the strongest ones in control cells, and were mostly associated with promoters. It is possible that the appearance of new origins upon incubation with PhenDC3 reduced the need of very strong origins.

New origins represented 71% of all origins affected by PhenDC3-mediated G4 stabilization, and showed a level of activity similar to that of *insensitive* origins. The appearance of these new OGRE/G4-containing origins might compensate the suppression of origins that lack OGRE/G4.

Examination of specific loci during *X. laevis* early development has shown that initiation of DNA replication did not require specific sites<sup>53,54</sup>, in contrast with late development, when site-specific initiation of DNA replication correlates with transcription onset in the embryo<sup>55</sup>. This regulation was explained by the huge excess of replication factors in *X. laevis* eggs, and by the short cell cycle (30 min) without G1 and G2 phases during the first 12 cell cycles after fertilization. Here, we found that OGRE/G4 oligonucleotides, but not random or AT-rich oligonucleotides, are strong competitors for replication origin activity in this system. We showed that this competition is at the level of DNA replication initiation and not at the level of complementary DNA strand synthesis. The pre-incubation with OGRE/G4 oligonucleotides did not affect pre-RC formation on origins, but only DNA synthesis activation. This suggests that some factors involved in this process are sequestered by the competing OGRE/G4 oligonucleotides. Our results might suggest a new explanation to the rapid replication cycles of *Xenopus* early embryos. Indeed, it is now recognized that potential origins are in large excess relative to those effectively activated in a given cell. The inter-origin spacing in a somatic cell is around 100 kb. If all origins were to be activated in a given cell, this spacing would be less than 10 kb. A full usage of specific origins would be therefore compatible with the speed of DNA replication in *X. laevis* early development.

How could G4 structures be involved in DNA replication initiation? From *E. coli* to higher eukaryotes, origins usually contain an origin recognition site, where the pre-RC is assembled, upstream of the initiation site of DNA synthesis, where nascent DNA strands are initiated by the DNA polymerase machinery. The origin recognition site may play a regulatory role, similar to transcription promoters that are localized 50 to 300 bp upstream of the TSS. An important feature of the OGRE/G4 element is its localization not at the initiation site of DNA synthesis, but 250 bp upstream of it<sup>2–4</sup>, suggesting an interaction with factors involved in the pre-RC. Figure 7 illustrate this position and show that our present data also confirm this position. This localization would fit with the site of assembly of the preRC, in agreement with the observation that recombinant ORC preferentially binds to G4 sequences<sup>8</sup>. However, alternatively OGRE/G4 elements could be part of sequences that regulate DNA synthesis initiation, possibly explaining the present discrepancy between its role in origin recognition and its replication fork stalling activity<sup>56</sup>. It is worth noting that our oligonucleotide experiments in *Xenopus* egg extracts point out to a role in the activation of DNA replication origins rather than in the assembly of the preRC. Known factors involved in this activation step are the kinase activity (DDK) which phosphorylates MCMs subunits and a complex



**Fig. 7** G4 function in the DNA replication initiation. Based on the position of the OGRE/G4 that we already reported<sup>2</sup> and the corresponding model<sup>3</sup>, we know that the OGRE/G4 element is always upstream of the initiation site itself, either on the + or the - DNA strand (left and right upper panels). In the corresponding lower panels, based on our present data, we confirm again the position of the G4 element relative to the initiation site. The position of the G4 might fit with the position of the preRC, but this does not obligatorily imply it will recruit some factors to the pre-RC. First, G4 can adopt several types of structure which themselves might regulate the folding or the replication origin region. Second, G4 might play a role in the removal of the nucleosome positioned at the initiation site itself, a process necessary to load the replication machinery, and therefore regulate activation of DNA replication. Third, because G4 are themselves nucleosome-free regions, they could facilitate DNA helix opening at the initiation site

reaction engaging several activating factors such as Sld2, Sld3 and GINS, cdc 45, Dbp11. OGRE/G4 elements could play a role in this activation step by helping the recruitment of these factors. G4 sequences can adopt several different G4 structures, which possibly may have different roles in the processing of the preRC to the activation step. Another important feature of OGRE/G4 element is that they are nucleosome-free, in contrast to the initiation site itself which contains a positioned nucleosome<sup>4</sup>. In such, OGRE/G4 may also recruit factors involved in the removal of the positioned nucleosome during the activation step of DNA replication, therefore facilitating the recruitment of the DNA polymerase machinery and its associated factors. From an evolutionary point of view, the use of structural elements, such as a G4-forming sequence, to set the replication program might be advantageous because it is not dependent on strict sequence specificity. As these elements are widely present in the genome, their function could be less affected by potential point mutations than strict consensus sequences.

## Methods

**Cell culture.** CGR8 mouse ES cells (obtained from Austin Smith's laboratory, Department of Biochemistry University of Cambridge, UK) were cultured on gelatin-coated dishes (feeder-free, to avoid DNA contamination by mouse embryonic fibroblasts) in Glasgow Minimum Essential Medium (GMEM) supplemented with 2 mM glutamine, 0.05 mM 2-mercaptoethanol, 1000 units/ml Leukemia Inhibitory Factor (LIF) and 10% Fetal Bovine Serum (FBS). To study the effect of G4 stabilization on origin firing, cells were grown in the presence of 0.5% DMSO or 10  $\mu$ M PhenDC3 (in 0.5% DMSO). NIH3T3 cells (NIH/3T3 (ATCC CRL-1658) were grown in Dulbecco's modified Eagle's minimal (DMEM) medium supplemented with 10% FBS.

## Genetic modification using the CRISP/Cas9 technology

**Surveyor assay.** The gRNAs for targeted Cas9-driven genetic modifications were designed using the ZiFiT Targeter Software Version 4.2 (<http://zifit.partners.org/ZettoniFIT/Disclaimer.aspx>). The specificity of the designed gRNAs was tested in the Surveyor assay using the T7 endonuclease (ref NEB #E3321) with the primers SURV\_C\_S697, SURV\_C\_AS697 (for sequences see Supplementary Table 2). Successful modification of the chosen region was confirmed by gel electrophoresis of the obtained products (Supplementary Fig. 2B).

**Ectopic origin creation and deletion experiments.** Ectopic origin creation in mouse NIH 3T3 cells was obtained by lipofectamine (Invitrogen, ref. 18324-012) transfection of the MLM3639 plasmid expressing the Cas9 endonuclease (<https://www.addgene.org/42252/>), MLM3639 plasmid expressing a gRNA specific to the targeted region (gRNA insertion F, gRNA insertion R) (<https://www.addgene.org/43860/>), linearized pBluescript plasmid bearing the template for homologous recombination, and pBABE-puro vector encoding the puromycin resistance gene (<https://www.addgene.org/34589/>). Cells were selected in medium containing 2.5  $\mu$ g/ml puromycin. The insertion presence was confirmed using the C3 AS1, ori1 G1, A5 S2, and ori1 C1 primers (Supplementary Table 2), and the absence of random insertions of the linearized pBluescript plasmid using the primers pBS1529S and pBS1726AS (Supplementary Table 2). Clones positive for homologous recombination were amplified for nascent strand purification.

**OGRE/G4 deletion from an endogenous origin.** Deletion of an OGRE/G4 from an endogenous replication origin was obtained by transfection of the MLM3639 plasmid expressing the Cas9 nickase (hCas9\_D10A) (<https://www.addgene.org/41816/>), two different MLM3639 plasmids to express gRNAs specific to the targeted regions (gRNA Ori1 delG4 1 F, gRNA Ori1 delG4 1 R, gRNA Ori1 delG4 2 F, gRNA Ori1 delG4 2 R; for sequences see Supplementary Table 2), and the pBABE-puro vector encoding the puromycin resistance gene. After puromycin selection, cells were cloned and checked for the presence of mutations using the MslI restrictase that recognizes a specific sequence in the vicinity of the targeted region (for experimental outcome see Supplementary Fig. 3A). The region of interest was amplified from clones bearing mutations using the primers Ori1 742 F and Ori1 742 R (Supplementary Table 2) and subcloned in pBluescript for precise mutation mapping by sequencing.

**RNA-primed short nascent strand (SNS) DNA strand isolation.** SNS were purified as described in<sup>4</sup> and in Supplementary Methods. The Illumina TruSeq ChIP Sample Prep Set A (ref 15034288) was used for preparation of sequencing libraries. Samples were sequenced using the Illumina HiSeq 2000 at the MGX Genomix facility (Montpellier). To perform local origin mapping, purified nascent strand samples were amplified by qPCR using the specific primers listed in Supplementary Table 2 with the LightCycler 480 SYBR Green Master mix (Roche, ref. 04887352001) on a LightCycler 480 II apparatus (Roche). The nascent strand enrichment was calculated as the ratio of the signal scored at origin-specific and background regions. If not otherwise specified, the statistical analysis was performed with the two-tailed, unpaired *t* test and the enrichment detected in 3 independent experiments. Differences with *p* values  $\leq 0.05$  were considered as statistically significant.

**Local transcription activity measurement.** Total cell RNA was extracted using the RNeasy Mini Kit (ref 74104 Qiagen) and cDNA was synthesized using the First-Strand cDNA Synthesis Kit with SuperScript II and a polyA primer (Invitrogen), according to the manufacturer's protocol. The transcription activity of selected genes was measured by qPCR with specific primers designed at the exon-intron junctions to avoid amplification from any possible DNA contaminant (Rai1c4ex3-4, Rai1 qPCR130, Gapdh ex4-5, Actb-ex2-3, Actb; see Supplementary Table 2). The relative transcription level was calculated as the transcription level found in the mutated versus parental cell line. The mean  $\pm$  SD was calculated from three independent experiments and the statistical evaluation was performed with the two-tailed, paired *t* test (*p* value  $\leq$  0.05 was considered significant).

**Read mapping.** Sequenced reads were mapped against the mm10 mouse genome sequence (NCBI GRCm38) using Bowtie2. Origins identification was obtained using MACS2 (version 2.1.0, ref. <sup>57</sup> (narrow peaks) and SICER (broad region). MACS2 peaks overlapping SICER regions were considered as actual replication initiation sites (IS). Three biological replicates of control mouse ES cells incubated with 0.5% DMSO and two replicates of mouse ES cells incubated with 10  $\mu$ M PhenDC3 were used as well as one RNase A-treated sample prior to  $\lambda$  exonuclease digestion (control). Only origins reproducibly present in at least two replicates in each condition were retained for further analysis. For figures representing raw data (UCSC tracks Fig. 3c, and Heatmap Fig. 3d), the mapped reads from replicates incubated with DMSO or PhenDC3 were merged for simplicity. Differential binding analysis was performed using the DESeq2 option in the DiffBind R package (version 1.12.3). The resulting *p* values were subjected to Benjamini-Hochberg multiple testing correction to derive the false discovery rates (FDR); only sites differentially bound with a FDR  $\leq$  1% were considered as differential. As a negative control for peak clustering, correlation with chromatin marks and motif discovery, the shuffle program from the Bedtools suite (v2.25.0<sup>29</sup>) was used to select random genomic regions of the same number and sizes as the origin peaks.

**Genomic localization.** Origin localized at promoters (2 kb upstream TSS) in transcribed and intergenic regions were identified using the GenomicRanges R package and the TxDb.Mmusculus.UCSC.mm10.knownGene, version 3.0.0, genome database. For negative controls, the IS coordinates were shuffled 1000 times while keeping the chromosomal distribution of each class and avoiding long regions lacking genomic information.

**G4 assignation.** Putative G4 were identified using the G4-Hunter algorithm<sup>26</sup> and a score higher than 2. An IS was considered as G4-positive if the G4 (with a G4Hunter\_score  $\geq$  2) was located  $\pm$ 500 bp from its center. The G4-Hunter score evaluates the propensity of a sequence to form a G4. A sequence with a G4-Hunter score higher than 2 should form a G4; to date, no sequence with such score was unable to form a G4 in classical experimental conditions (37 °C, neutral pH, 100 mM NaCl or KCl).

**G4 profile.** G4 location profiles were computed by counting the "G4 location" at the base pair level at  $\pm$ 1 kb from the IS for each origin class. Then, the sum of the coverage, or the G4 ratio for each group, was computed to obtain the G4 profiles for each origin class. Profiles of G4 on the minus strand (CCC) were oriented on the (+) strand.

**RNA-seq and differential gene expression.** Total RNA was extracted using the RNeasy Mini Kit (Qiagen; cat 74104), and libraries were prepared using the Illumina TruSeq Stranded mRNA Sample Preparation Kit and sequenced using an Illumina HiSeq 2500 apparatus at the MGX GenomiX facility (Montpellier). The TopHat software (version 2.1.1) was used for splice junction mapping with Bowtie2 (version 2.2.9) for mapping reads. Reads counting on genes was done using HTSeq-count (version 0.6.1p1). Data were normalized to the relative log expression implemented in edgeR (version 3.16.5), and the statistical analysis to identify differentially expressed genes was performed using DESeq2 (version 1.14.1). Differential gene expression was considered when the adjusted *p* value was  $\leq$  0.05 after multi-testing correction using the Benjamini-Hochberg method.

**Genomic Set Enrichment Analysis (GSEA).** The GSEA was performed using the R package fgsea (version 1.2.1) and the data obtained in the differential RNA-seq analysis. Genes were ranked from upregulated to downregulated using the adjusted *p* value and the sign of the fold change obtained from the DESeq2 analysis. The enrichment set test was computed with the genes associated with one of the origin classes (suppressed, new, etc.), and the *p* value was computed using 10,000 permutations (origin-gene associations).

**De novo motif discovery.** The RSAT peak-motifs program<sup>58</sup> was used to detect de novo motifs around the IS summits from  $-1$  kb to  $+1$  kb. Among the results, the motif found by positions-analysis for 6-7 nt and with the lowest *e*-value and the highest significance was selected.

**Episomal DNA replication assay.** The HEK-293 cell line that stably expressing EBNA1 (HEK293 EBNA1+) was cultured in DMEM with 10% fetal calf serum and 220  $\mu$ g/ml neomycin. The HEK293 cells was originally received from DSMZ (DSMZ No: ACC 305). CMV-EBNA1 was stably integrated into the chromosome after linearization and selected with 220  $\mu$ g/ml Neomycin. Episomal replication was assayed using the Dpn1 digestion method<sup>59</sup>. The reporter plasmids (2  $\mu$ g) containing the various origin variants were transfected in HEK293 cells that express EBNA1, and the transfection efficiencies were verified by visualizing GFP-positive cells. Six days post-transfection, cells were harvested using the protocol described by Hirt et al.<sup>60</sup> Isolated DNA was purified by phenol-chloroform extraction and digested with 40 U DpnI (NEB) in the presence of RNase (Roche). Digested DNA (300 ng) was electroporated in Electromax DH10B competent cells (Invitrogen) and ampicillin-resistant colonies, representing the number of recovered plasmids, were counted to estimate the replication efficiency of the episome (presented as the mean  $\pm$  standard deviation calculated from 3 to 7 independent experiments). Statistical significance was evaluated using the two-tailed, unpaired *t* test (*p* value  $\leq$  0.05 was considered as significant).

**X. laevis egg extract and DNA replication kinetics.** Low Speed Egg (LSE) and High-Speed Egg extracts (HSE) were prepared as previously described<sup>61,62</sup> (for details see Supplementary Methods). Chromosomal DNA replication was assayed by adding demembrated *X. laevis* sperm nuclei to extracts supplemented with [ $\alpha$ -<sup>32</sup>P]-dCTP. For competition assays, extracts were incubated with 2 ng/ $\mu$ l of oligonucleotides (or shared salmon sperm as control, or ultrapure water) at 22 °C for 5 min before sperm nuclei addition. DNA synthesis was monitored by TCA precipitation. Incorporated acid-insoluble material was spotted onto Whatman glass microfiber filters, grade GF/C, and then precipitated with 5% TCA solution containing 2% pyrophosphate. After ethanol washes, filters were dried and the incorporated TCA-precipitated radioactivity was counted in scintillation liquid. M13 replication kinetics were assessed using 400 ng of ssDNA per 50  $\mu$ l of HSE<sup>62</sup> pre-incubated or not with oligonucleotides Sperm chromatin purification for protein-binding monitoring was performed as previously described<sup>61</sup>. Briefly, chromatin pellets were resuspended in 2 $\times$  LB (0.125 M Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 10% 2- $\beta$ -mercaptoethanol and 0.004% bromophenol blue), denatured at 95 °C for 5 min, and then stored at  $-20$  °C or immediately analyzed by SDS-PAGE, using gradient Bis-Tris gels (Thermo Fisher Scientific).

**Antibodies.** The antibodies used in this work were against: H3 (Abcam, ab1791, dilution 1/2000), H2B (Abcam, ab1790, dilution 1/2000), phosphorylated CHK1 (Cell Signaling, 2341 S, dilution 1/250), PCNA (Sigma, P8825, dilution 1/2500), RPA34<sup>62</sup> (dilution 1/500), MCM3<sup>61</sup> (dilution 1/2000), CDC45<sup>63</sup> (dilution 1/1000), ELYS<sup>31,64</sup> (dilution 1/500), MCM4<sup>63</sup> (dilution 1/1000), anti-Chk1 (dilution 1/500), anti-ORC5 (dilution 1/1000), anti-CDC6<sup>63</sup> (dilution 1/500), OCT4 (Abcam, ab19857, dilution 1/500), actin (Sigma, A4700, dilution 1/500), HRP-linked ECL anti-mouse IgG (GE Healthcare, NA931V, dilution 1/4000), HRP-linked ECL anti-rabbit IgG (GE Healthcare, NA934V, dilution 1/4000) (For details see Supplementary Table 7).

**Spectroscopic studies.** Isothermal difference spectra (IDS) and circular dichroism (CD) measurements were performed as previously described<sup>17,65</sup>. Briefly, the sequences were tested at 4  $\mu$ M strand concentration in 10 mM LiCaco pH 7.2 with 100 mM KCl. IDS were obtained by computing the difference between the absorbance spectra of unfolded and folded oligonucleotides that were recorded before and after addition of 100 mM KCl, respectively, at 25 °C. CD spectra were recorded at 20 °C after IDS (in K+) on a JASCO-1500 spectropolarimeter using 1 cm path length quartz cuvettes.

**FRET melting assay and FRET competition assay.** The tested G4 sequences (Table S3) were labeled with Fam on 5' and Tamra on 3'. Each sequence was pre-folded at 0.2  $\mu$ M in 10 mM LiCaco pH 7.2 supplemented with 10 mM KCl and 90 mM LiCl before adding the PhenDC3 ligand (1  $\mu$ M). Stabilization (increase in  $T_{1/2}$ , expressed in °C) was plotted for each G4-forming sequence; as a control a dsDNA of the same length were used. In the FRET competition assay, stabilization ( $\Delta T_{1/2}$ , in °C) of the human telomeric quadruplex F21T by 0.5  $\mu$ M PhenDC3 was analyzed in the presence/absence of increasing amounts of each G-rich origin sequence (3 or 10  $\mu$ M strand concentration).

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

## Data availability

The SNS-seq and RNA-seq data are deposited at the NCBI GEO (GSE126477) [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE126477>] and [[http://rsat-tagc.univ-mrs.fr/g4/g4\\_data.html](http://rsat-tagc.univ-mrs.fr/g4/g4_data.html)]. R scripts used for figure creation are deposited under [[https://github.com/LacroixLaurent/G4Hunter\\_mm10\\_Orj](https://github.com/LacroixLaurent/G4Hunter_mm10_Orj)] and [http://rsat-tagc.univ-mrs.fr/g4/g4\\_data.html](http://rsat-tagc.univ-mrs.fr/g4/g4_data.html). Data supporting the findings of this study are available within the paper and its supplementary information files, including uncropped scans of the most important blots. All the data are available from the authors upon reasonable request.

**Code availability**

For MACS2 see <https://github.com/taoliu/MACS>. For SICER see <https://home.gwu.edu/~wpeng/Software.htm>. For GenomicRanges: <https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html>. For DESeq2 see <http://bioconductor.org/packages/release/bioc/html/DESeq2.html>. For G4-Hunter see <https://github.com/LacroixLaurent/G4HunterPaperGit>. For Fgsea <https://bioconductor.org/packages/release/bioc/html/fgsea.html>. For RSAT [http://rsat-tagc.univ-mrs.fr/rsat/RSAT\\_home](http://rsat-tagc.univ-mrs.fr/rsat/RSAT_home). A reporting summary for this Article is available as a Supplementary Information file.

Received: 13 November 2018 Accepted: 8 May 2019.

Published online: 22 July 2019

**References**

- Fragkos, M., Ganier, O., Coulombe, P. & Mechali, M. DNA replication origin activation in space and time. *Nat. Rev. Mol. Cell Biol.* **16**, 360–374 (2015).
- Cayrou, C. et al. Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res.* **21**, 1438–1449 (2011).
- Cayrou, C. et al. New insights into replication origin characteristics in metazoans. *Cell Cycle* **11**, 658–667 (2012).
- Cayrou, C. et al. The chromatin environment shapes DNA replication origin organization and defines origin classes. *Genome Res.* **25**, 1873–1885 (2015).
- Comoglio, F. et al. High-resolution profiling of Drosophila replication start sites reveals a DNA shape and chromatin signature of metazoan origins. *Cell Rep.* **11**, 821–834 (2015).
- Langley, A. R., Graf, S., Smith, J. C. & Krude, T. Genome-wide identification and characterisation of human DNA replication origins by initiation site sequencing (ini-seq). *Nucleic Acids Res.* **44**, 10230–10247 (2016).
- Valton, A. L. et al. G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *EMBO J.* **33**, 732–746 (2014).
- Hoshina, S. et al. Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. *J. Biol. Chem.* **288**, 30161–30171 (2013).
- Kumagai, A. & Dunphy, W. G. MTBP, the partner of Treslin, contains a novel DNA-binding domain that is essential for proper initiation of DNA replication. *Mol. Biol. Cell* **28**, 2998–3012 (2017).
- Sugimoto, N., Maehara, K., Yoshida, K., Ohkawa, Y. & Fujita, M. Genome-wide analysis of the spatiotemporal regulation of firing and dormant replication origins in human cells. *Nucleic Acids Res.* **46**, 6683–6696 (2018).
- Costas, C. et al. Genome-wide mapping of Arabidopsis thaliana origins of DNA replication and their associated epigenetic marks. *Nat. Struct. Mol. Biol.* **18**, 395–400 (2011).
- Rodriguez-Martinez, M. et al. The gastrula transition reorganizes replication-origin selection in Caenorhabditis elegans. *Nat. Struct. Mol. Biol.* **24**, 290–299 (2017).
- Cadoret, J. C. et al. Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proc. Natl Acad. Sci. USA* **105**, 15837–15842 (2008).
- Bartholdy, B., Mukhopadhyay, R., Lajugie, J., Aladjem, M. I. & Bouhassira, E. E. Allele-specific analysis of DNA replication origins in mammalian cells. *Nat. Commun.* **6**, 7051 (2015).
- Smith, O. K. et al. Distinct epigenetic features of differentiation-regulated replication origins. *Epigenetics Chromatin* **9**, 18 (2016).
- Vorlickova, M. et al. Circular dichroism and guanine quadruplexes. *Methods* **57**, 64–75 (2012).
- Mergny, J. L., Li, J., Lacroix, L., Amrane, S. & Chaires, J. B. Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.* **33**, e138 (2005).
- Lindner, S. E., Zeller, K., Schepers, A. & Sugden, B. The affinity of EBNA1 for its origin of DNA synthesis is a determinant of the origin's replicative efficiency. *J. Virol.* **82**, 5693–5702 (2008).
- Hammerschmidt, W. & Sugden, B. Replication of Epstein-Barr viral DNA. *Cold Spring Harb. Perspect. Biol.* **5**, a013029 (2013).
- Chaudhuri, B., Xu, H., Todorov, I., Dutta, A. & Yates, J. L. Human DNA replication initiation factors, ORC and MCM, associate with oriP of Epstein-Barr virus. *Proc. Natl Acad. Sci. USA* **98**, 10085–10089 (2001).
- Dhar, S. K. et al. Replication from oriP of Epstein-Barr virus requires human ORC and is inhibited by geminin. *Cell* **106**, 287–296 (2001).
- Schepers, A. et al. Human origin recognition complex binds to the region of the latent origin of DNA replication of Epstein-Barr virus. *EMBO J.* **20**, 4588–4602 (2001).
- De Cian, A., Delemos, E., Mergny, J. L., Teulade-Fichou, M. P. & Monchaud, D. Highly efficient G-quadruplex recognition by bisquinolinium compounds. *J. Am. Chem. Soc.* **129**, 1856–1857 (2007).
- Chung, W. J., Heddi, B., Hamon, F., Teulade-Fichou, M. P. & Phan, A. T. Solution structure of a G-quadruplex bound to the bisquinolinium compound Phen-DC(3). *Angew. Chem. Int. Ed. Engl.* **53**, 999–1002 (2014).
- Halder, R., Riou, J. F., Teulade-Fichou, M. P., Frickey, T. & Hartig, J. S. Bisquinolinium compounds induce quadruplex-specific transcriptome changes in HeLa S3 cell lines. *BMC Res. Notes* **5**, 138 (2012).
- Bedrat, A., Lacroix, L. & Mergny, J. L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **44**, 1746–1759 (2016).
- Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33**, 2908–2916 (2005).
- Martin, M. M. et al. Genome-wide depletion of replication initiation events in highly transcribed regions. *Genome Res.* **21**, 1822–1832 (2011).
- Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
- Lunyak, V. V., Ezrokhi, M., Smith, H. S. & Gerbi, S. A. Developmental changes in the Sciarra II/9A initiation zone for DNA replication. *Mol. Cell Biol.* **22**, 8426–8437 (2002).
- Blow, J. J., & Laskey, R. A. Initiation of DNA replication in nuclei and purified DNA by a cell-free extract of Xenopus eggs. *Cell* **47**, 577–587 (1986).
- Mechali, M. & Harland, R. M. DNA synthesis in a cell-free system from Xenopus eggs: priming and elongation on single-stranded DNA in vitro. *Cell* **30**, 93–101 (1982).
- Anderson, J. A., Lewellyn, A. L. & Maller, J. L. Ionizing radiation induces apoptosis and elevates cyclin A1-Cdk2 activity before but not after the midblastula transition in Xenopus. *Mol. Biol. Cell* **8**, 1195–1206 (1997).
- Kappas, N. C., Savage, P., Chen, K. C., Walls, A. T. & Sible, J. C. Dissection of the XChk1 signaling pathway in Xenopus laevis embryos. *Mol. Biol. Cell* **11**, 3101–3108 (2000).
- Kumagai, A. & Dunphy, W. G. Claspin, a novel protein required for the activation of Chk1 during a DNA replication checkpoint response in Xenopus egg extracts. *Mol. Cell* **6**, 839–849 (2000).
- Gillespie, P. J., Khoudoli, G. A., Stewart, G., Swedlow, J. R. & Blow, J. J. ELYS/MEL-28 chromatin association coordinates nuclear pore complex assembly and replication licensing. *Curr. Biol.* **17**, 1657–1662 (2007).
- Jares, P. & Blow, J. J. Xenopus cdc7 function is dependent on licensing but not on XORC, XCdc6, or CDK activity and is required for XCdc45 loading. *Genes Dev.* **14**, 1528–1540 (2000).
- Sequeira-Mendes, J. et al. Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS Genet.* **5**, e1000446 (2009).
- Delgado, S., Gomez, M., Bird, A. & Antequera, F. Initiation of DNA replication at CpG islands in mammalian chromosomes. *EMBO J.* **17**, 2426–2435 (1998).
- Mukhopadhyay, R. et al. Allele-specific genome-wide profiling in human primary erythroblasts reveal replication program organization. *PLoS Genet.* **10**, e1004319 (2014).
- Cayrou, C., Gregoire, D., Coulombe, P., Danis, E. & Mechali, M. Genome-scale identification of active DNA replication origins. *Methods* **57**, 158–164 (2012).
- Lipford, J. R. & Bell, S. P. Nucleosomes positioned by ORC facilitate the initiation of DNA replication. *Mol. Cell* **7**, 21–30 (2001).
- Belsky, J. A., MacAlpine, H. K., Lubelsky, Y., Hartemink, A. J. & MacAlpine, D. M. Genome-wide chromatin footprinting reveals changes in replication origin architecture induced by pre-RC assembly. *Genes Dev.* **29**, 212–224 (2015).
- Eaton, M. L., Galani, K., Kang, S., Bell, S. P. & MacAlpine, D. M. Conserved nucleosome positioning defines replication origins. *Genes Dev.* **24**, 748–753 (2010).
- Xu, J. et al. Genome-wide identification and characterization of replication origins by deep sequencing. *Genome Biol.* **13**, R27 (2012).
- Masai, H. et al. Molecular architecture of G-quadruplex structures generated on duplex Rif1 binding sequences. *J. Biol. Chem.* **293**, 17033–17049 (2018).
- Madireddy, A. et al. G-quadruplex-interacting compounds alter latent DNA replication and episomal persistence of KSHV. *Nucleic Acids Res.* **44**, 3675–3694 (2016).
- Norseen, J., Johnson, F. B. & Lieberman, P. M. Role for G-quadruplex RNA binding by Epstein-Barr virus nuclear antigen 1 in DNA replication and metaphase chromosome attachment. *J. Virol.* **83**, 10336–10346 (2009).
- Zamiri, B., Mirceta, M., Bomsztyk, K., Macgregor, R. B. Jr. & Pearson, C. E. Quadruplex formation by both G-rich and C-rich DNA strands of the C9orf72 (GGGGCC)<sub>8</sub>\*(GGCCCC)<sub>8</sub> repeat: effect of CpG methylation. *Nucleic Acids Res.* **43**, 10055–10064 (2015).
- Tsukakoshi, K., Saito, S., Yoshida, W., Goto, S. & Ikebukuro, K. CpG methylation changes G-quadruplex structures derived from gene promoters and interaction with VEGF and SP1. *Molecules* **23**, E944(2018).
- Mao, S. Q. et al. DNA G-quadruplex structures mold the DNA methylome. *Nat. Struct. Mol. Biol.* **25**, 951–957 (2018).

52. Gomez, M. & Brockdorff, N. Heterochromatin on the inactive X chromosome delays replication timing without affecting origin usage. *Proc. Natl Acad. Sci. USA* **101**, 6923–6928 (2004).
53. Laskey, R. A. & Harland, R. M. Replication origins in the eucaryotic chromosome. *Cell* **24**, 283–284 (1981).
54. Mechali, M. & Kearsley, S. Lack of specific sequence requirement for DNA replication in *Xenopus* eggs compared with high sequence specificity in yeast. *Cell* **38**, 55–64 (1984).
55. Hyrien, O., Maric, C. & Mechali, M. Transition in specification of embryonic metazoan DNA replication origins. *Science* **270**, 994–997 (1995).
56. Tubbs, A. et al. Dual roles of poly(dA:dT) tracts in replication initiation and fork collapse. *Cell* **174**, 1127–1142 e19 (2018).
57. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
58. Medina-Rivera, A. et al. RSAT 2015: regulatory sequence analysis tools. *Nucleic Acids Res.* **43**, W50–W56 (2015).
59. Gerhardt, J., Jafar, S., Spindler, M. P., Ott, E. & Schepers, A. Identification of new human origins of DNA replication by an origin-trapping assay. *Mol. Cell Biol.* **26**, 7731–7746 (2006).
60. Hirt, B. Selective extraction of polyoma DNA from infected mouse cell cultures. *J. Mol. Biol.* **26**, 365–369 (1967).
61. Maiorano, D., Cuvier, O., Danis, E. & Mechali, M. MCM8 is an MCM2-7-related protein that functions as a DNA helicase during replication elongation and not initiation. *Cell* **120**, 315–328 (2005).
62. Francon, P. et al. A hypophosphorylated form of RPA34 is a specific component of pre-replication centers. *J. Cell Sci.* **117**, 4909–4920 (2004).
63. Lemaitre, J. M., Bocquet, S. & Mechali, M. Competence to replicate in the unfertilized egg is conferred by Cdc6 during meiotic maturation. *Nature* **419**, 718–722 (2002).
64. Franz, C. et al. MEL-28/ELYS is required for the recruitment of nucleoporins to chromatin and postmitotic nuclear pore complex assembly. *EMBO Rep.* **8**, 165–172 (2007).
65. Guedin, A., Alberti, P. & Mergny, J. L. Stability of intramolecular quadruplexes: sequence effects in the central loop. *Nucleic Acids Res.* **37**, 5559–5567 (2009).

## Acknowledgements

We would like to thank Jacques van Helden (Aix-Marseille University) and our lab members for helpful discussions. We thank Stéphane Bocquet (IGH) for help in *Xenopus* egg extracts, J. Walter for the anti-CDC45 antibodies, I. Mattaj and J. Blow for the anti-ELYS antibody. We are grateful to the Genotoul Bioinformatics Platform Toulouse Midi-Pyrenees for computing and storage resources. We also thank E. Andermarcher for critical reading of the manuscript. The research leading to these results has received funding from the European Research Council (FP7/2007–2013 Grant Agreement no.233339). This work was also supported by the Association Française contre les

Myopathies (AFM), the ARC foundation and ANR14-CE10-0019, and by the MSDA-VENIR Fund GENE-IGH. PP was supported by a post-doctoral fellowship from the ARC Foundation (Fondation ARC pour la Recherche sur le Cancer).

## Author contributions

M.A., P.C. and A.A. contributed equally. M.M. proposed the project and the experimental system, P.P. designed and performed the majority of experiments, P.P. and P.C. designed the genome-modification experiments, L.L., M.A. and B.B. performed the bioinformatics analysis, A.A. performed the experiment with *X. laevis* egg extracts, I.P. cultured the cells and did FACS analysis, A.G. and J.L.M. designed and analysed in vitro G4-formation assays, J.D. and A.S. performed episome replication assay, P.P. and M.M. wrote and revised the manuscript. C.C. proposed the experimental system and the design for the experiments involving Phen DC, and M.P.F.T. provided PhenDC3 and helped in the use of this drug. All the authors read and approved the final manuscript.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-019-11104-0>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Peer review information:** *Nature Communications* thanks Hans Joachim Lipps and other anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

ARTICLE

<https://doi.org/10.1038/s41467-019-08905-8>

OPEN

# G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma

Yuxiang Wang<sup>1</sup>, Jie Yang<sup>2</sup>, Aaron T. Wild<sup>1</sup>, Wei H. Wu<sup>1</sup>, Rachna Shah <sup>1</sup>, Carla Danussi<sup>3</sup>, Gregory J. Riggins<sup>4</sup>, Kasthuri Kannan<sup>5</sup>, Erik P. Sulman <sup>2,6</sup>, Timothy A. Chan <sup>1,7</sup> & Jason T. Huse<sup>3,6</sup>

Mutational inactivation of *ATRX* ( $\alpha$ -thalassemia mental retardation X-linked) represents a defining molecular alteration in large subsets of malignant glioma. Yet the pathogenic consequences of *ATRX* deficiency remain unclear, as do tractable mechanisms for its therapeutic targeting. Here we report that *ATRX* loss in isogenic glioma model systems induces replication stress and DNA damage by way of G-quadruplex (G4) DNA secondary structure. Moreover, these effects are associated with the acquisition of disease-relevant copy number alterations over time. We then demonstrate, both in vitro and in vivo, that *ATRX* deficiency selectively enhances DNA damage and cell death following chemical G4 stabilization. Finally, we show that G4 stabilization synergizes with other DNA-damaging therapies, including ionizing radiation, in the *ATRX*-deficient context. Our findings reveal novel pathogenic mechanisms driven by *ATRX* deficiency in glioma, while also pointing to tangible strategies for drug development.

<sup>1</sup>Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA. <sup>2</sup>Department of Radiation Oncology, University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. <sup>3</sup>Department of Translational Molecular Pathology, University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. <sup>4</sup>Departments of Neurosurgery, Oncology, and Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD 21231, USA. <sup>5</sup>Department of Pathology, New York University School of Medicine, New York, NY 10016, USA. <sup>6</sup>Department of Pathology, University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. <sup>7</sup>Department of Radiation Oncology, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA. Correspondence and requests for materials should be addressed to J.T.H. (email: [jhuse@mdanderson.org](mailto:jhuse@mdanderson.org))

Infiltrating gliomas are the most common primary brain tumors and, despite considerable molecular and clinical heterogeneity, remain uniformly deadly in the face of aggressive surgical and cytotoxic treatment regimens<sup>1</sup>. Recent large-scale genomic profiling has shown that inactivating mutations in *ATRX* ( $\alpha$ -thalassaemia mental retardation X-linked) characterize large subclasses of both adult and pediatric glioma<sup>2–4</sup>. Despite these striking correlations, however, the precise mechanisms by which *ATRX* mutation promotes gliomagenesis remain unclear. Recent reports have linked germline *ATRX* mutations to osteosarcoma<sup>5–7</sup>, and their association with a rare, congenital neurodevelopmental condition (*ATR-X* syndrome) is well-established<sup>8</sup>. *ATRX* encodes a chromatin binding protein widely implicated in epigenetic regulation and remodeling<sup>9–15</sup>, suggesting that epigenomic dysfunction may, at least in part, underlie the oncogenic effects of *ATRX* deficiency. *ATRX* loss has also been implicated in alternative lengthening of telomeres (ALT), an abnormal telomerase-independent mechanism of telomere maintenance based on homologous recombination<sup>16,17</sup>. Finally, *ATRX* deficiency has been repeatedly linked to replication stress, DNA damage, copy number alterations (CNAs), and aneuploidy<sup>18–23</sup>, and recent work has associated *ATRX* deficiency specifically with copy number loss at ribosomal DNA loci<sup>24</sup>. Whether and how such genomic instability contributes to the initiation and/or evolution of malignant glioma remains unclear.

*ATRX* binds widely across the genome at sites featuring tandem repeats and CpG islands<sup>25</sup>. Many such loci are GC-rich and susceptible to forming G-quadruplexes (G4s), abnormal secondary structures implicated in both transcriptional dysregulation and DNA damage. Accordingly, it has been hypothesized that, among its various functionalities, *ATRX* serves to resolve G4s genome-wide and mitigate their deleterious consequences<sup>25,26</sup>. The tendency of G4s to stall replication forks underlies their association with DNA damage<sup>27</sup>. Chemical stabilization of G4s induces replication stress at genomic loci prone to G4 formation<sup>28</sup>, and also promotes DNA damage and apoptosis in neural progenitor cells<sup>29</sup>. Moreover, recent work suggests that G4-induced replication stress at telomeres may drive ALT in the *ATRX*-deficient setting through induction of homologous recombination<sup>16</sup>. Indeed, G4 stabilization hampers the ability of forced *ATRX* expression to abrogate the ALT phenotype in vitro. Taken together, these findings provide compelling links between *ATRX*, G4 biology, and genomic instability. Whether *ATRX* deficiency directly induces G4 formation and DNA damage, however, remains unestablished, as does the impact of G4s on the pathogenesis of *ATRX*-deficient neoplasia. Moreover, therapeutic strategies leveraging G4 biology in the selective targeting of *ATRX*-deficient cancers have not been extensively explored.

To characterize the role of G4-mediated genomic instability in glioma biology, we inactivated *ATRX* in isogenic normal human astrocyte (NHA) and glioma stem cell (GSC) models. We found that *ATRX* loss increased G4 formation, replication stress, and DNA damage genome-wide. Moreover, *ATRX*-deficient NHAs accumulated clinically relevant CNAs at an accelerated rate relative to *ATRX*-intact counterparts. Chemical G4 stabilization was associated with enhanced DNA damage and cell death in *ATRX*-deficient contexts. Moreover, *ATRX*-mutant GSC xenografts were selectively sensitive to G4-targeting in vivo. Finally, G4 stabilization in *ATRX*-deficient NHAs and GSCs effectively synergized with other DNA-damaging treatment strategies, including ionizing radiation. These findings clarify distinct mechanisms by which G4s influence *ATRX*-deficient glioma pathogenesis and indicate that G4 stabilization may represent an attractive therapeutic strategy for the selective targeting of *ATRX*-mutant cancers.

## Results

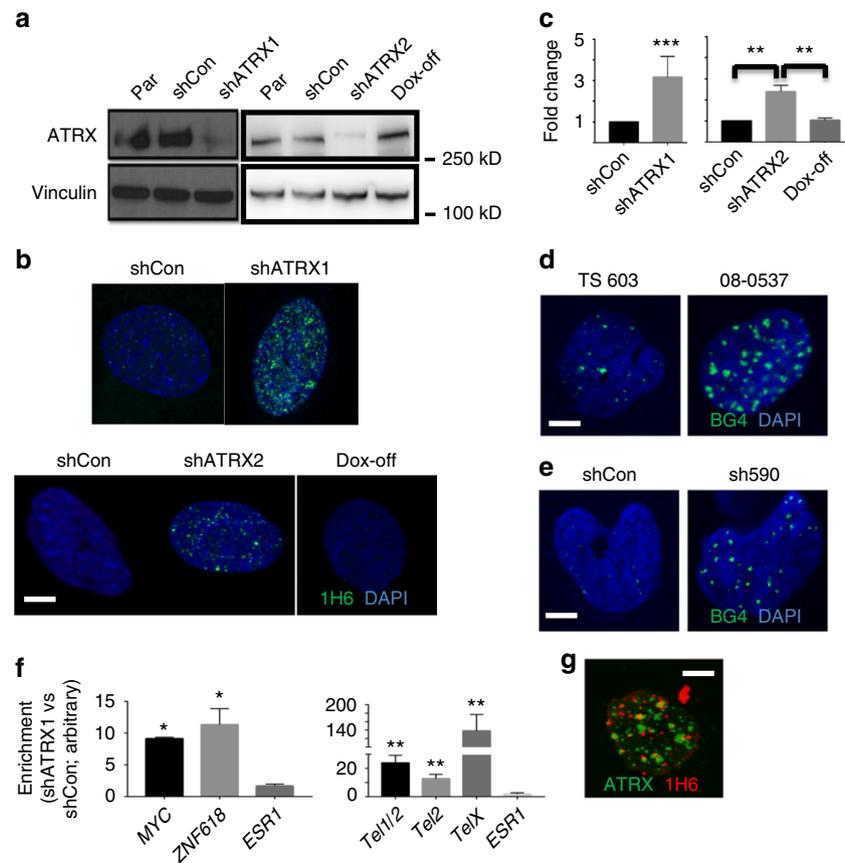
### ***ATRX* deficiency promotes G4 formation and DNA damage.**

To model the genomic consequences of *ATRX* deficiency in a glioma-relevant cellular context, we performed shRNA-mediated *ATRX* knockdown in TERT and E6/E7-transformed NHAs. Several studies have effectively employed immortalized NHAs to delineate key aspects of glioma biology<sup>30–34</sup>. In our investigations, we employed two distinct hairpin constructs to silence *ATRX*—sh*ATRX1* and sh*ATRX2*—the latter of which was driven by a doxycycline (dox)-inducible promoter (Fig. 1a). This framework allowed for the analysis of both immediate and long-term effects of *ATRX* deficiency as well as their reversibility. Using a monoclonal antibody known to recognize G4 structures in situ (1H6), we then demonstrated that *ATRX* deficiency increased nuclear G4s relative to levels seen in control shRNA-expressing parental NHAs (shCon), an effect that was reversible upon restored *ATRX* expression (Fig. 1b–c). Increased G4s were also found in p53-deficient murine neuroepithelial progenitor cells (mNPCs) featuring inactivated *Atrx* (Supplementary Fig. 1a).

The specificity of 1H6 for DNA-based secondary structures was confirmed by DNAase treatment, which eradicated immunolabelling, and RNase treatment, which did not, in NHAs treated with the G4-stabilizing agent CX-3543 (see below, Supplementary Fig. 1b). Moreover, the effects of *ATRX* knockdown on nuclear G4 levels in isogenic NHAs, as assessed by 1H6 immunofluorescence, were recapitulated with a different G4-targeting monoclonal antibody (BG4; Supplementary Fig. 1c). Forced expression of the isocitrate dehydrogenase 1 (IDH1) R132H mutation in our isogenic NHAs did not significantly alter G4 levels (as assessed by BG4) in either the *ATRX*-intact or the *ATRX*-deficient context (Supplementary Fig. 1d). *ATRX* deficiency almost invariably co-occurs with mutations in *IDH1* or its homologue *IDH2* in adult gliomas. We also compared GSCs derived from IDH-mutant, *ATRX*-mutant (08-0537) and IDH-mutant, *ATRX* wild-type (TS 603) gliomas, finding increased nuclear G4s in the former by BG4 immunofluorescence (Fig. 1d). Finally, we found that *ATRX* knockdown (sh590) in a glioblastoma-derived GSC (TS 543; IDH and *ATRX* wild type) enhanced G4 formation (Fig. 1e and Supplementary Fig. 1e). These data show that, in multiple glioma-relevant cellular contexts, *ATRX* deficiency promotes G4 accumulation.

To further support these findings, we employed a synthetic single-chain antibody (hf2) to immunoprecipitate G4s in both *ATRX*-intact and *ATRX*-deficient contexts. hf2 specificity was validated by gel-shift assay showing specific capture of synthesized Kit2 nucleotides independently from random ssDNA and dsDNA (Supplementary Fig. 2a). We then performed pulldowns in our isogenic NHAs, finding that *ATRX* deficiency significantly increased the qPCR enrichment of known G4 sites within the *MYC* and *ZNF618* loci, as well as in telomeric regions on chromosomes 1, 2, and X (Fig. 1f)<sup>35–37</sup>. Consistent with the notion that *ATRX* resolves G4s as part of its normal functionality, we found a distinct absence of colocalization between *ATRX* and G4 immunofluorescence in *ATRX*-intact NHAs (Fig. 1g). Finally, functional studies demonstrated that *ATRX* knockdown failed to induce significant changes in apoptosis, BrdU incorporation, or cell cycle profile (Supplementary Fig. 2b–d). Taken together, these findings confirm, in a true isogenic system, that *ATRX* deficiency promotes G4 formation. Moreover, they indicate that, at least in this glioma-relevant context, increased G4s as a consequence of *ATRX* deficiency are insufficient to drive apoptosis or impact cellular proliferation.

We then examined whether the G4s induced by *ATRX* deficiency promoted replication stress and DNA damage, as suggested by prior literature<sup>38</sup>. We found that *ATRX* knockdown significantly and reversibly increased  $\gamma$ -H2AX and



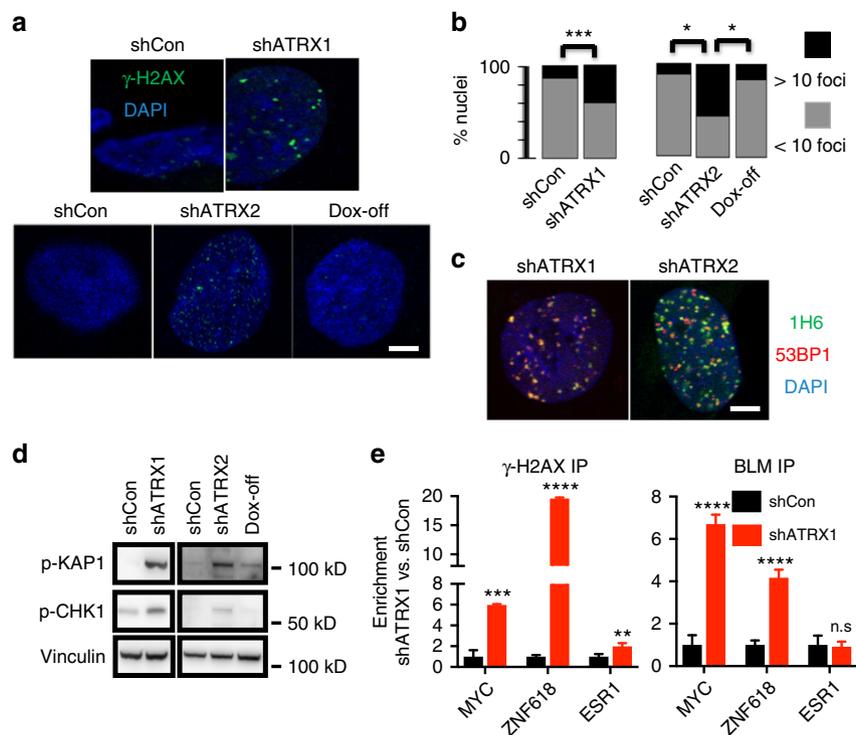
**Fig. 1** ATRX deficiency promotes G4 formation. **a** Western blots for ATRX in parental (Par), shControl (shCon), and shATRX NHA (Vinculin loading control). Left panel shows constitutive NHA lines (shATRX1) and right panel shows the inducible lines post Doxycycline induction (shATRX2) and withdrawal (Dox-off). **b** Immunofluorescent staining of G4 (1H6) in constitutive (upper panel) and inducible (lower panel) NHA lines (DAPI counterstain). **c** Quantified relative G4 signal intensity (50 nuclei counted in all cases). **d, e** Immunofluorescent staining of G4 (BG4) in ATRX-intact (TS 603) and ATRX-mutant (08-0537) GSCs (**d**) and in TS 543 GSCs (**e**) subjected to ATRX knockdown (sh590) versus control (shCon; DAPI counterstain). **f** G4-containing DNA fragments from shCon and shATRX1 NHAs (three replicates each) were pulled down from sheared genomic DNA with purified hf2 antibodies. Recovered DNA was subjected to real-time PCR for telomeric sequences or G4-rich promoter regions of *MYC* and *ZNF618* (*ESR1* used as negative control). Graphs show enrichment over GAPDH scaled to shCon levels. **g** ATRX and G4 immunofluorescence in parental NHAs showing no significant colocalization. Where applicable, error bars reflect SEM; *P* values determined by unpaired, two-tailed *t*-test (\**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001); scale bars represent 10  $\mu$ m

53BP1-positive DNA damage foci by immunofluorescence (Fig. 2a–c), and did so in a pattern that extensively colocalized with nuclear G4 distribution, as assessed by both 1H6 and BG4 (Fig. 2c and Supplementary Fig. 3). Moreover, these changes were accompanied by engagement of the replication stress pathway, as evidenced by upregulated levels p-CHK1 and p-KAP1 on western blot (Fig. 2d). To further ascertain the extent of colocalization between G4s and DNA damage sites in the ATRX-deficient context, we performed chromatin immunoprecipitation (ChIP) for  $\gamma$ -H2AX and Bloom syndrome RecQ-like helicase (BLM), a protein known to bind G4 DNA<sup>39</sup>. We found that ATRX loss significantly increased both  $\gamma$ -H2AX and BLM enrichment at putative G4 sites within the *MYC* and *ZNF618* loci (Fig. 2e), far exceeding effects at a negative control locus (*ESR1*). To determine whether increased levels of DNA damage in the ATRX-deficient setting might lead to structural abnormalities in chromosomes, we performed metaphase cytogenetic analysis coupled with telomere-fluorescence in situ hybridization (FISH) in shCon and shATRX1 NHAs uniformly aged to 15 passages. Consistent with multiple prior reports, ATRX knockdown in this context was not associated with an obvious ALT phenotype (Supplementary Fig. 4a, b). However, we consistently observed

higher levels of chromosome breakage in ATRX-deficient NHAs relative to shCon counterparts (Fig. 3a and Supplementary Fig. 4b). Increased chromosome breaks were also observed in ATRX-mutant 08-0537 GSCs relative to ATRX wild-type TS 603 GSCs (Fig. 3b and Supplementary Fig. 4c), and a similar trend, though not statistically significant, was seen in p53-deficient mNPCs also featuring *Atrx* inactivation (versus p53-deficient, *Atrx*-intact isogenics; Fig. 3c and Supplementary Fig. 4d). These data establish pathogenic links between G4s arising with ATRX deficiency and the generalized genomic instability characteristic of ATRX-mutant tumors and cell lines.

#### ATRX deficiency drives clinically relevant CNA formation.

Having confirmed that ATRX deficiency induces DNA damage and structural abnormalities in chromosomes, likely through G4-mediated mechanisms, we sought to assess whether these biological processes might promote acquisition of CNAs in ATRX-deficient tumors. ATRX mutations in adult glioma arise almost exclusively in the setting of concurrent mutations in *TP53* and either *IDH1* or *IDH2*. The glioma subtype defined by this combined genotype, termed “IDHmut-noncode”<sup>2</sup>, features



**Fig. 2** ATRX deficiency promotes replication stress and DNA damage. **a**  $\gamma$ -H2AX immunofluorescence in constitutive (upper panel) and inducible (lower panel) NHAs (DAPI counterstain). **b** Quantified percentage of cells with >10 immunopositive  $\gamma$ -H2AX foci (50 nuclei counted in all cases). **c** NHAs with ATRX knockdown were double stained for G4 and 53BP1, showing extensive colocalization. **d** Western blots of p-KAP1 and p-CHK1 show activation of replication stress arising with ATRX knockdown. **e**  $\gamma$ -H2AX and BLM ChIP in shCon and shATR1 NHAs (three replicates each) showing that ATRX deficiency promotes enrichment for both proteins at putative G4-rich promoter regions (MYC and ZNF618) far exceeding that seen at a negative control site (ESR1). Where applicable, error bars reflect SEM; *P* values determined by unpaired, two-tailed *t*-test (\**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001, \*\*\*\**P* < 0.0001, n.s.: *P* > 0.05); scale bars represent 10  $\mu$ m

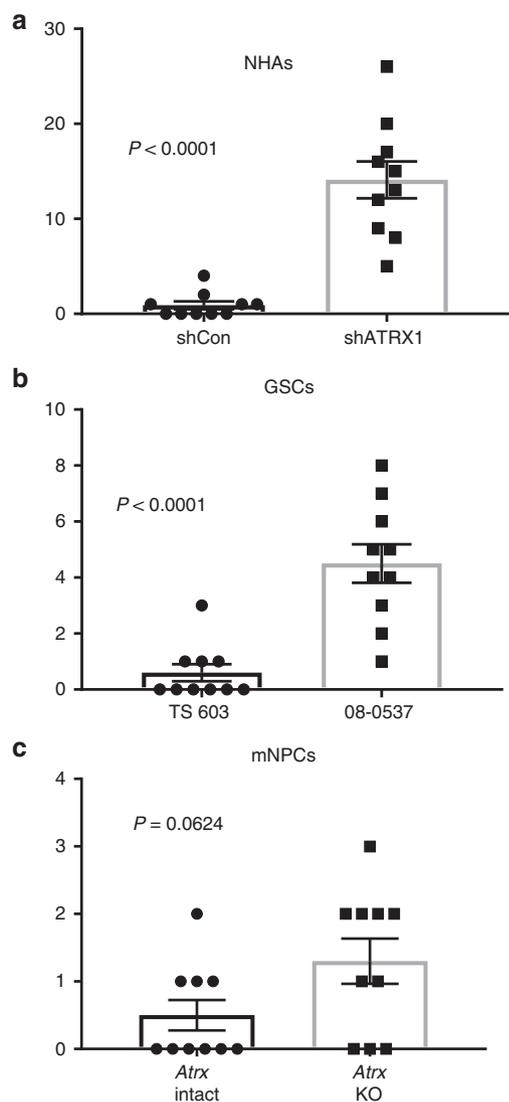
almost uniformly low-level *ATRX* expression and exhibits a characteristic pattern of CNAs, distinct from that commonly seen in other adult glioma subtypes (Supplementary Fig. 5)<sup>2</sup>. Moreover, multiple CNAs recurrently featured in ATRX-deficient glioma mobilize established oncogenic and/or tumor suppressive loci, including *MYC* and *CDKN2A*<sup>2</sup>, implying that such structural abnormalities may contribute to the malignant evolution of this inexorably progressive cancer.

To experimentally model CNA formation in the ATRX-deficient setting, we aged our isogenic NHAs in culture, monitoring DNA copy number by SNP array at 5 and 15 passages—~1 month and ~3 months in culture, respectively (Supplementary Fig. 5). We found that while both sets of isogenics demonstrated increased CNAs over time, ATRX-deficient NHAs exhibited a distinct pattern of gains and losses that included larger (>1 Mb) alterations not seen in ATRX-intact counterparts (Fig. 4a). Analysis of TCGA SNP data revealed a similar subset of broad alterations included within the CNA profile of the IDHmut-noncodel glioma subtype (Fig. 4b). Moreover, two of the broad CNAs arising with ATRX deficiency in NHAs, involving 12p gain and 14q loss, recapitulated events commonly seen in the IDHmut-noncodel glioma subtype and associated with relatively unfavorable prognosis when present (Fig. 4c–f). Taken together, these findings support the notion that G4-mediated DNA damage induces specific patterns of CNAs in the ATRX-deficient, IDHmut-noncodel glioma subtype, which in turn influence malignant evolution.

**Chemical G4 stabilization selectively targets ATRX-deficient cells.** As indicated above, the pronounced effects of ATRX

deficiency on G4 formation and replication stress in NHAs were not associated with increased cell death at baseline. Nevertheless, we reasoned that compensatory mechanisms to resolve G4s and otherwise maintain genomic integrity were likely under increased stress, and that chemical stabilization of G4s might, therefore, selectively enhance DNA damage to an unsustainable degree in the ATRX-deficient context. To evaluate the therapeutic potential of this synthetic lethal approach, we treated our isogenic NHAs in culture with increasing concentrations of CX-3543 (Quarflorin), an established G4-stabilizing agent<sup>40,41</sup>. We found that ATRX knockdown, in both constitutive and inducible systems, was associated with increased sensitivity to CX-3543 (IC<sub>50</sub> = 42.449 nM (shATR1) vs 328.835 nM (shCon) in constitutive NHAs and IC<sub>50</sub> = 357.424 nM (pre-induction) vs 53.415 nM (shATR2) vs 247.700 nM (post-induction) in inducible NHAs; Fig. 5a, b). Similar results were obtained with two other G4-stabilizing agents, pyridostatin (PDS) and CX-5461 (Supplementary Fig. 6a–b)<sup>28,42</sup>. Clonogenicity studies also revealed enhanced vulnerability to CX-3543 in ATRX-deficient NHAs as well as TS 543 GSCs subjected to ATRX knockdown (sh590; Fig. 5c, d). Restoring ATRX expression reverted NHAs to baseline levels of sensitivity (Fig. 5b).

$\gamma$ -H2AX immunofluorescence demonstrated dramatically increased levels of DNA damage in ATRX-deficient NHAs treated with CX-3543, accompanied by activation of the replication stress pathway as determined by western blot (Fig. 5e–h). Similar effects on  $\gamma$ -H2AX immunofluorescence were observed using either PDS or CX-5461 in NHAs (Supplementary Fig. 7a), and were also seen in both ATRX-knockdown TS 543 and ATRX-mutant 08-0537 GSCs treated with CX-3543, each



**Fig. 3** ATX deficiency induces chromosome breaks. **a** ATX-deficient NHAs (shATRX1, passage 15) showed significantly increased chromosome breaks by cytogenetic analysis relative to ATX-intact controls (shCon). **b** ATX-mutant GSCs (08-0537) showed significantly increased chromosome breaks by cytogenetic analysis relative to ATX wild-type GSCs (TS 603). **c** Quantified chromosome breaks in *Atrx*-intact and *Atrx*-KO mNPCs (also *Tp53*<sup>-/-</sup>, passage 10). In all cases, 10 sets of chromosomes were quantified. Error bars reflect SEM; *P* values determined by unpaired, two-tailed *t*-test

relative to ATX-intact counterparts (Supplementary Fig. 7b). Moreover, 53BP1-positive DNA damage foci arising with CX-3543 demonstrated extensive colocalization with G4s on confocal microscopy (Supplementary Fig. 8a–b), a finding recapitulated by  $\gamma$ -H2AX immunofluorescence in ATX-deficient NHAs treated with either CX-3543, PDS, or CX-5461, using either BLM or BG4 immunofluorescence to designate G4s (Supplementary Fig. 8c–d). Once again, these effects were reversed following ATX re-expression (Fig. 5f, h, and Supplementary Fig. 8b). Finally, G4 stabilization in ATX-deficient NHAs bolstered the extent of ChIP enrichment for both BLM and  $\gamma$ -H2AX at putative G4 sites (Supplementary Fig. 9a–9b), further demonstrating that causal links between impaired G4 resolution and DNA damage underlie our synthetic lethal approach. Annexin V flow-cytometry confirmed that the heightened sensitivity of

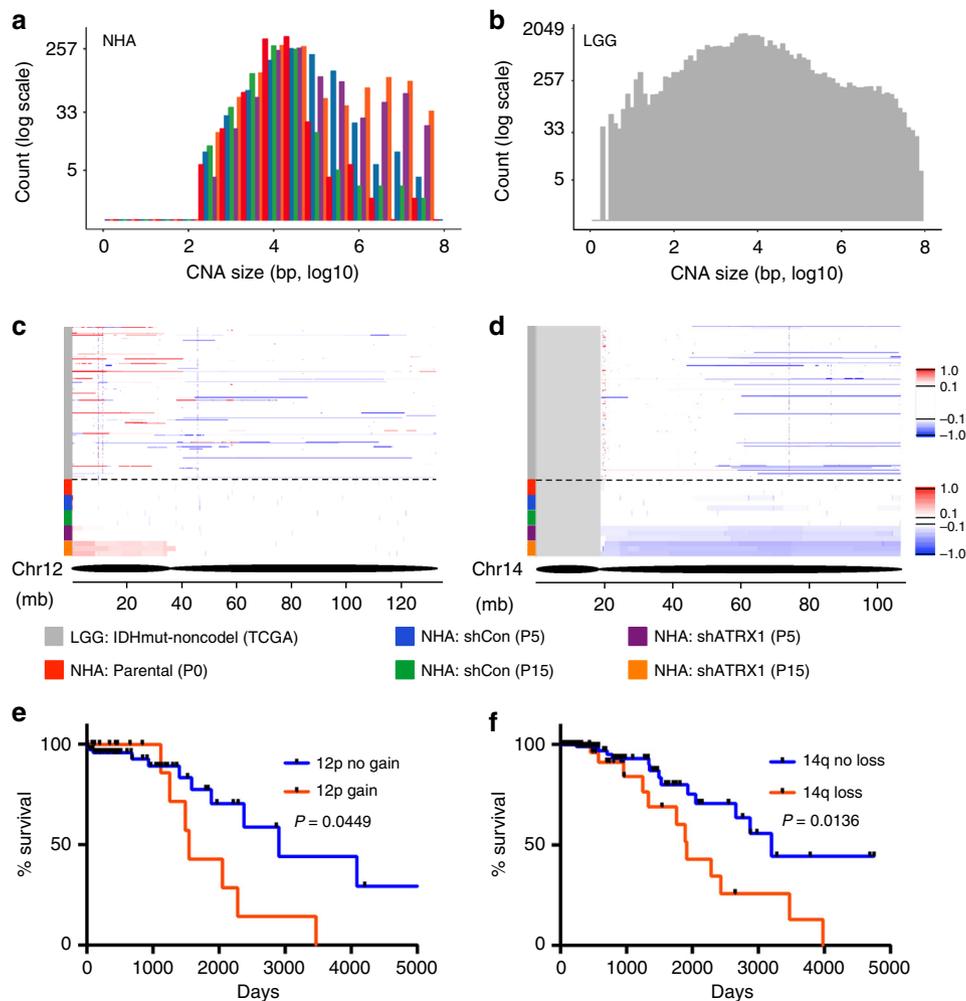
ATX-deficient NHAs to CX-3543 reflected increased apoptosis, and this enhanced cell death followed the kinetics of replication stress pathway activation in both constitutive and inducible isogenic contexts (Fig. 6a–d). Analogous experiments in ATX-intact and ATX-knockdown TS 543 GSCs yielded similar findings (Supplementary Fig. 10a–b). In total, these results indicate that chemical stabilization of G4 structures selectively promotes cell death in the ATX-deficient context, likely by inducing toxic levels of DNA damage.

The experimental links, described above, between replication stress, DNA damage, and CX-3543 treatment prompted us to consider whether G4 stabilization might enhance the therapeutic efficacy of established DNA-damaging treatment strategies, particularly in ATX-deficient context. To evaluate this possibility, we subjected vehicle and CX-3543-treated isogenic NHAs, cultured in soft agar, to increasing doses of either ionizing radiation (IR) or hydroxyurea (HU), assessing viable colonies at 21 days. We found that CX-3543 treatment potentiated the cytotoxicity of both IR and HU, and while these effects were significant for both NHA genotypes, they were particularly strong in the setting of ATX deficiency (Fig. 6e, f). Restoring ATX expression markedly dampened the extent of cytotoxic synergy (Fig. 6e, f). Moreover, the ATX-dependent radiosensitization properties of CX-3543 were recapitulated in TS 543 GSC isogenics, (Supplementary Fig. 10c). These findings inform additional therapeutically relevant strategies combining chemical G4 stabilization with well-established treatment modalities in the targeting of ATX-deficient cancer.

#### ATX loss enhances sensitivity to G4 stabilization in vivo.

Having established the increased sensitivity of ATX-deficient NHAs and GSCs to chemical G4 stabilization in cell culture, we sought to ascertain whether this approach could exhibit a similar degree of efficacy in vivo. To this end, we employed an ATX-mutant, patient-derived GSC line (JHH-273) capable of forming tumors in murine hosts when embedded in the hind flank<sup>43</sup>. Following cellular implantation, we subjected xenografted mice to daily intravenous treatment with either CX-3543 or vehicle and monitored tumor growth over time. We found that CX-3543 dramatically slowed the growth of JHH-273 flank tumors (Fig. 7a, b and Supplementary Fig. 11a) and significantly prolonged survival in xenografted mice (Fig. 7c). Histopathological examination of CX-3543-treated xenografts revealed cellular depopulation, reduced proliferative activity by Ki-67 immunostaining, and increased  $\gamma$ -H2AX-positive DNA damage foci relative to untreated counterparts, recapitulating in vitro findings (Fig. 7d). Notably, telomere FISH failed to reveal changes in the level of ALT in residual viable tumor following CX-3543 treatment (Fig. 7d).

To ascertain whether these effects were specific to the ATX-mutant context, we performed analogous xenograft experiments using ATX wild-type TS 543 cells. In these studies, we found that CX-3543 treatment had no appreciable effect on either xenograft growth or mouse survival (Fig. 8a, b and Supplementary Fig. 11b). However, when we subjected these same GSCs to ATX knockdown, they were rendered sensitive to CX-3543 to an extent similar to that observed for JHH-273 cells (Fig. 8c, d and Supplementary Fig. 11c). ATX knockdown also recapitulated histopathological effects on cellular depopulation, proliferative activity, and  $\gamma$ -H2AX-positivity (Fig. 8e, f). Consistent with prior reports<sup>19,22</sup>, ATX knockdown was not associated with ALT in TS 543 cells (Fig. 8e, f). Taken together, these in vivo findings further support the therapeutic potential for chemical G4 stabilization in the selective targeting of ATX-deficient glioma.



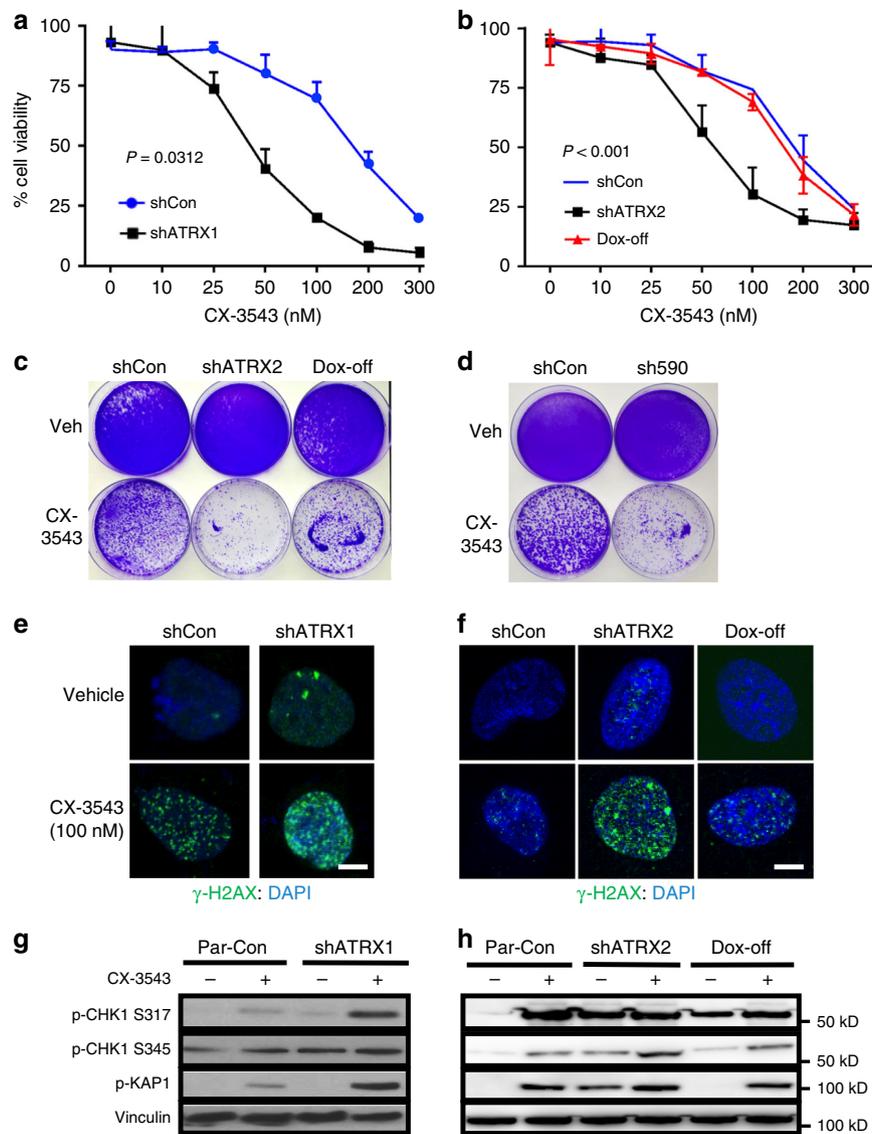
**Fig. 4** ATRX deficiency induces clinically relevant copy number alterations. **a** Size distribution of CNAs in parental (P0), shCon (P5 and P15), and shATRX1 (P5 and P15) NHAs (three replicates each). In ATRX knockdown cells, large CNAs (>1 Mb) arise over passage number. **b** Size distribution of CNAs in IDHmut-noncode gliomas (TCGA)<sup>2</sup>. **c, d** IGV plots for chromosome 12 (**c**) and 14 (**d**) comparing CNA regions in IDHmut-noncode gliomas (above dotted line) to NHAs (below dotted line). Color scales indicate log<sub>2</sub> copy number. **e, f** Kaplan-Meier curves for IDHmut-noncode glioma patients with or without either 12p gain or 14q loss, showing significant differences in overall survival. *P* values determined by Log-rank (Mantel-Cox) test

## Discussion

As indicated above, loss-of-function mutations in *ATRX* likely play central pathogenic roles in several distinct tumor variants, including multiple subtypes of incurable glioma. That *ATRX* itself encodes a chromatin regulatory protein suggests that epigenomic dysfunction underlies, at least in part, the oncogenic sequelae of its inactivation. To this point, we recently demonstrated that *ATRX* deficiency induces extensive chromatin remodeling and transcriptional shifts in putative glioma cells of origin, driving disease-relevant phenotypes that modulate both cellular motility and differentiation<sup>44</sup>. However, the full impact of *ATRX* deficiency on tumor initiation and evolution almost certainly includes other molecular mechanisms. The association of *ATRX* mutation and ALT<sup>17,22</sup>, for instance, is now extensively described and provides a vehicle to telomerase-independent immortalization in affected cancer cells. Moreover, recent work has linked ALT to the well-characterized genomic instability induced by *ATRX* deficiency<sup>16,45</sup>.

The pathogenic consequences of *ATRX*-dependent genomic instability in the context of cancer are unknown. Abundant prior work has demonstrated links between *ATRX* deficiency, DNA damage, CNA development, and aneuploidy<sup>18–23</sup>. Indeed,

p53-dependent apoptosis derived from genomic instability in the neuroepithelial progenitor compartment likely underlies the neuronal depopulation, microcephaly, and mental retardation associated with ATR-X syndrome<sup>46</sup>. Replication stress has been extensively implicated as a root cause of genomic instability in *ATRX*-deficient cells<sup>19,21,47</sup>. In addition to activating DNA damage pathway signaling, replication fork stalling and collapse can generate double-strand breaks and defective chromosome condensation during mitosis, both of which are known to drive CNAs and aneuploidy of the kind seen in *ATRX*-mutant glioma<sup>48–50</sup>. Recent work strongly supports the notion that the replication stress characterizing *ATRX*-deficient cells derives, at least in part, from G4 DNA secondary structure<sup>16,27–29</sup>. *ATRX* binds widely at GC-rich genomic sites susceptible to forming G4s<sup>25</sup>, and restored *ATRX* expression in *ATRX*-mutant cell lines mitigates G4-associated phenotypes such as ALT<sup>16</sup>. Such data imply that *ATRX* may serve to protect the genome from unwanted G4 formation and the potentially deleterious consequences of ensuing genomic instability. Our findings support this conjecture by demonstrating, for the first time, that *ATRX* deficiency potently and reversibly induces G4 formation in isogenic experimental models ranging from transformed astrocytes



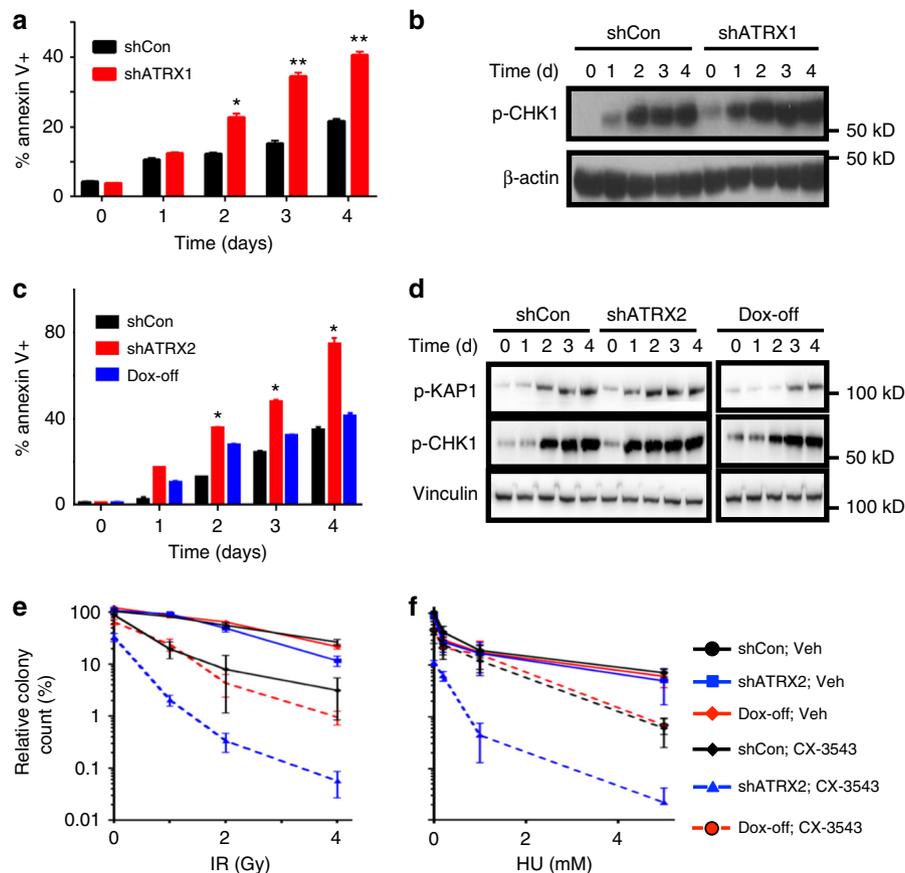
**Fig. 5** ATRX-deficient glioma models are selectively sensitive to G4 stabilization. **a, b** Cell viability (CellTiter-Glo) of constitutive (**a**) and inducible (**b**) shCon and shATRX NHAs (four replicates each) treated with CX-3543 from 0–300nM. **c** Clonogenic assay of inducible shATRX2 NHAs demonstrates enhanced and reversible sensitivity to CX-3543 (50 nM) with ATRX deficiency. **d** Clonogenic assay of TS 543 with (sh590) and without (shCon) ATRX knockdown demonstrates enhanced sensitivity to CX-3543 (50 nM) with ATRX deficiency. **e, f**  $\gamma$ -H2AX immunofluorescence in constitutive (**e**) and inducible (**f**) shATRX NHAs showing increased DNA damage with CX-3543 treatment (100 nM), particularly in the setting of ATRX knockdown. **g, h** Western blots showing increased phosphorylation of replication stress pathway constituents (CHK1 and KAP1) in constitutive (**g**) and inducible (**h**) shATRX NHAs following CX-3543 treatment (100 nM). Where applicable, error bars reflect SEM; P values determined by two-way ANOVA; scale bars represent 10  $\mu$ m

to patient-derived GSCs. As such, they confirm a novel functionality for a SWI/SNF epigenetic regulator already widely implicated in chromatin remodeling, structure, and organization.

That increased G4s were accompanied by replication stress signaling, DNA damage at spatially overlapping sites—as confirmed by both immunofluorescence and ChIP, and disease-relevant patterns of CNAs in our cell line models provides additional evidence that this pathobiological cascade features in ATRX-mutant neoplasia. Prior computational analysis across multiple tumor types established significant correlations between CNA breakpoints and genomic sites enriched in putative G4-forming sequences<sup>51</sup>, firmly implicating G4s in the process of cancer-associated CNA acquisition. In our NHA models, ATRX knockdown led to a distinct CNA profile over time enriched in alterations over 1 Mb in size. While this pattern did not

completely mirror the known CNA signature of ATRX-mutant gliomas<sup>2</sup>, it did recapitulate key elements involving larger, arm-level events. In particular, two CNAs (12p gain and 14q loss) were reminiscent of analogous alterations in human tumors associated with unfavorable prognosis. These data speak directly to the premise that CNA mobilization, driven at least in part by G4-mediated DNA damage, promotes malignant evolution in ATRX-deficient gliomas. As this tumor subtype characteristically progresses slowly over time<sup>52</sup>, such mechanistic insights are consistent with established clinical features.

Due to its sheer prevalence in glioma, ATRX deficiency represents a molecular target of intriguing therapeutic potential. That being said, effective strategies to drug an inactivated epigenetic regulator are not immediately obvious, as they might be in the setting of more conventional, kinase-predominant,



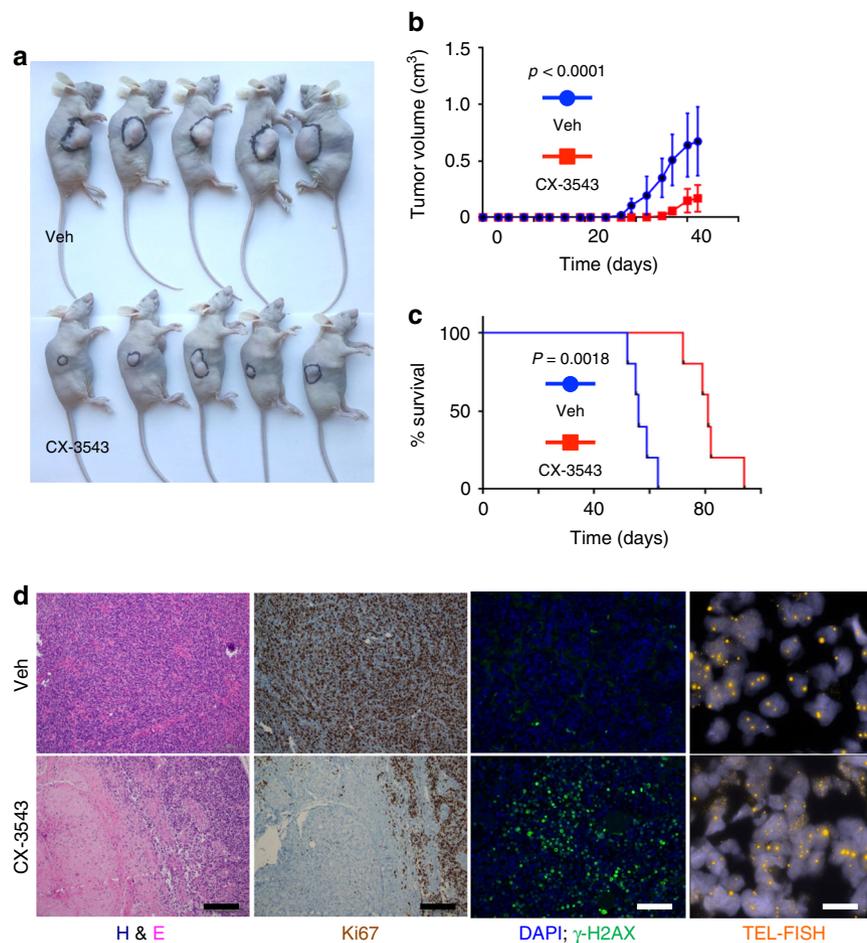
**Fig. 6** G4 stabilization selectively promotes apoptosis and synergizes with DNA-damaging therapies. **a, b** Time course study in shCon and shATRX1 NHAs (three replicates each) treated with 100 nM CX-3543 showing parallel kinetics of apoptosis (Annexin V positivity) (**a**) and p-CHK1/p-KAP1 levels by western blot (**b**). **c, d** Analogous time course study in inducible shATRX2 cells (three replicates each) documenting p-CHK1/p-KAP1 levels (**d**) by western blot and Annexin V positivity (**c**) over time. **e, f** Soft agar colony counts for NHAs (3 replicates each) treated with either vehicle control (Veh) or 50 nM CX-3543 and either increasing doses of IR (**e**) or HU (**f**). Colony counts measured at 21 days scaled to that of shCon, Veh NHAs. Where applicable, error bars reflect SEM; *P* values determined by two-way ANOVA (\**P* < 0.05, \*\**P* < 0.01)

oncogenic signaling networks. Given these challenges, leveraging specific vulnerabilities engendered by ATRX loss might offer alternative approaches. In particular, the longstanding association of ATRX deficiency with genomic instability, confirmed in this report, presents a tangible opportunity to explore synthetic lethality paradigms, akin to that of poly (ADP-ribose) polymerase (PARP) inhibitors in the treatment of *BRCA1*-inactivated breast cancer<sup>53</sup>. While the observed level of DNA damage in our ATRX-deficient cell line and tumor models was insufficient to induce apoptosis in isolation, due in part to coincident *TP53* inactivation, we hypothesized that its targeted enhancement would overwhelm compensatory mechanisms maintaining cell viability (Fig. 9). Moreover, our identification of G4s as the likely source of ATRX-deficient genomic instability provided a viable approach to therapeutic selectivity.

We found that G4 stabilization with multiple distinct agents selectively targeted ATRX-deficient glioma cell lines and tumors, both in vitro and in vivo. That our findings, initially obtained with CX-3543, were recapitulated with CX-5461, and PDS argues that on-target effects dependent on G4 binding were chiefly responsible for observed therapeutic impact. Moreover, cell death in these contexts was temporally associated with DNA damage and replication stress, further supporting impaired G4 resolution as a likely mechanism of action. These results recapitulate recent data showing enhanced sensitivity of ATRX-deficient embryonic stem cells to CX-5461<sup>24</sup>. We cannot completely exclude the

possibility that G4-stabilization exerts some of its cytotoxic effects through the manipulation of ALT. As alluded to above, prior work has functionally linked increased G4s and DNA damage at telomeres with ALT induction in ATRX-deficient cells<sup>16</sup>. Nevertheless, ATRX knockdown was not associated with ALT in our NHA and TS 543 isogenics, consistent with multiple prior reports<sup>19,20,22</sup>, and CX-3543 failed to alter the pattern of telomere FISH in ATRX-mutant JHH-273 GSC xenografts. Taken together, these findings strongly suggest that the cytotoxicity of G4 stabilization in the ATRX-deficient context is, at least in large part, mediated by DNA damage genome-wide, not limited to telomeric regions.

We also demonstrated that G4 stabilization dramatically enhanced the effects of IR and HU in ATRX-deficient NHAs, highlighting possibilities for effective synergistic combinations in the clinical setting. Since its introduction almost 40 years ago, IR has remained one of the most important nonsurgical therapeutic modalities employed in the treatment of malignant glioma, with demonstrated efficacy across disease subtypes<sup>54–56</sup>. Moreover, recent work has shown that ATRX-mutant gliomas in particular exhibit increased sensitivity to DNA-damaging combinations of IR and chemotherapy<sup>57,58</sup>. This vulnerability may derive in part from increased genomic instability at baseline. Defective non-homologous end joining (NHEJ), documented to arise with ATRX deficiency in preclinical models<sup>20</sup>, may also play a role. Regardless of the precise molecular mechanisms at



**Fig. 7** G4 stabilization markedly slows the growth of ATRX-mutant glioma xenografts. **a** representative image of mice bearing JHH-273 xenografts following treatment with either vehicle (veh) or 12.5 mg/kg CX-3543 for 42 days. **b** Average xenograft volume over time in 10 mice randomized (5 and 5) into two groups, receiving either 12.5 mg/kg CX-3543 or vehicle control (veh). **c** Kaplan-Meier analysis of survival in a separate cohort of 10 mice randomized (5 and 5) into two groups, receiving either 12.5 mg/kg CX-3543 or vehicle control (Veh). **d** Representative H&E-stained, immunostained (Ki67 and  $\gamma$ -H2AX), and TEL-FISH stained sections of xenografts from vehicle (Veh) and CX-3543-treated mice. CX-3543-treated tumors showed decreased cellularity, decreased proliferative activity, and increased DNA damage. *P* values were determined by two-way ANOVA; scale bars represent 200  $\mu$ m in staining panels and 25  $\mu$ m in Tel-FISH panels

work, therapeutically potentiating an already effective treatment strategy for glioma represents an underexplored approach with the potential for considerable clinical impact.

Precisely which G4 stabilizer represents the optimal agent for clinical translation remains unclear. Both CX-3543 and CX-5461 have advanced to clinical trials for pancreatic neuroendocrine tumor and BRCA1/2-deficient breast cancer, respectively<sup>40,42</sup>. However, CX-3543, by report, has limited bioavailability<sup>41</sup>, and no formal blood–brain barrier penetration studies have been released for either. Nevertheless, our findings indicate that the targeted approach of G4 stabilization has considerable therapeutic potential in the treatment of ATRX-deficient glioma, along with other ATRX-mutant cancers. That the strategy is based on a tumor-specific vulnerability arising in association with an easily assessable biomarker should facilitate its clinical application, while also minimizing harmful side effects in treated patients. Moreover, alternative G4-stabilizing agents are currently available for use both as tool compounds and starting points for chemical derivatization<sup>42,59,60</sup>.

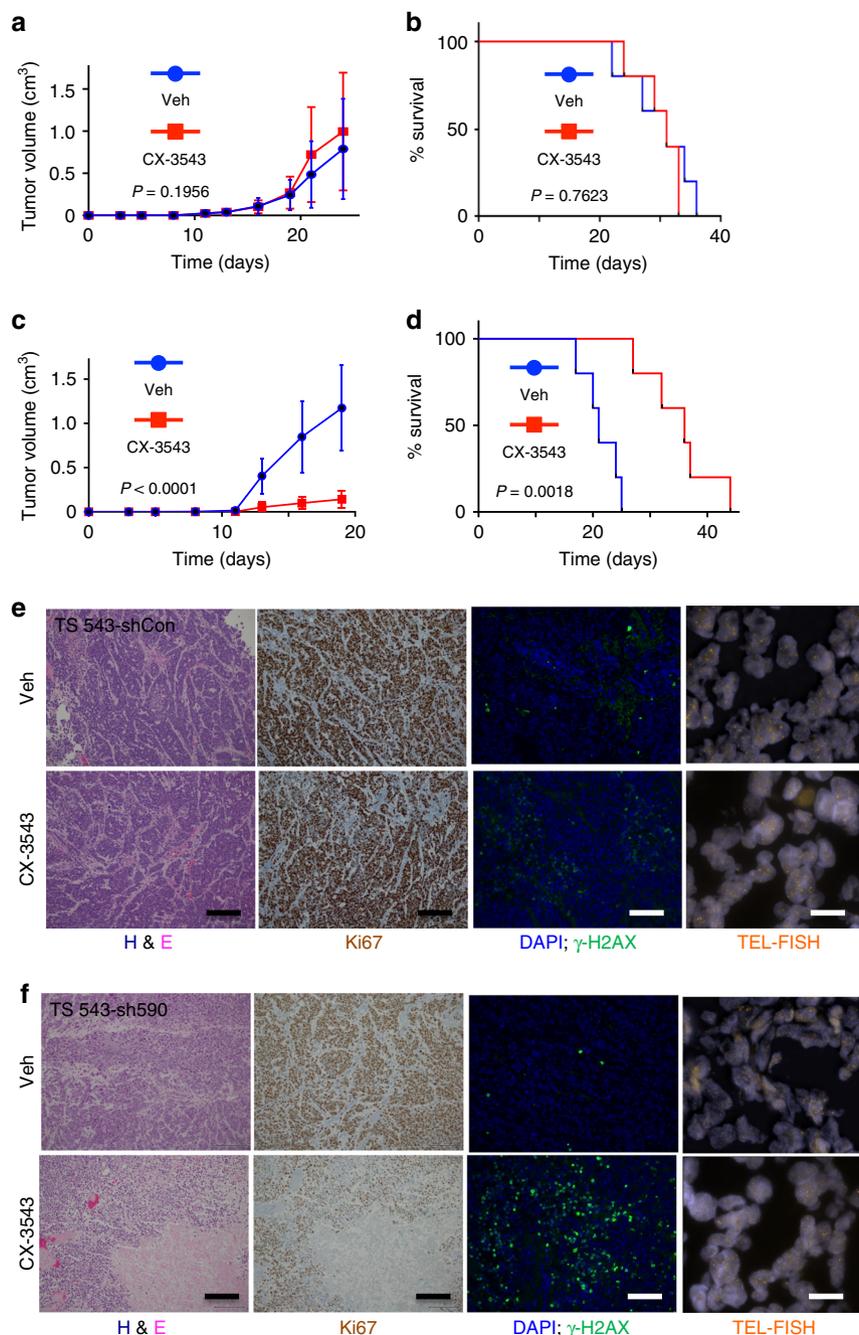
In summary, we firmly implicate G4 secondary structure as a defining characteristic of ATRX-mutant glioma, one that drives disease-relevant genomic instability and presents opportunities

for tangible therapeutic advancement. As such, our work has important implications for both the molecular pathogenesis of ATRX-deficient neoplasia, as well as the development of more effective drugs specifically targeting a palette of deadly tumors.

## Methods

**Study design.** The objective of this study was to determine the impact of ATRX deficiency on G4 formation, DNA damage, and genomic instability in glioma, and assess the potential of chemical G4 stabilization as a therapeutic strategy in ATRX-deficient tumors. This was a controlled, laboratory-based, experimental study using cell line models in culture and in xenografts. ATRX was inactivated by genetic approaches and, in some cases, pharmaceutical agents and/or ionizing radiation were applied. Sample sizes were determined independently for each experiment without formal power calculation. No data was excluded from analysis. Unless otherwise specified, experiments employed three replicates per sample. End points varied by experiment and are described below, in figure legends, or in the Results section. Histopathological and immunohistochemical review of xenografts was conducted by a Neuropathologist (J.T.H.) in a nonblinded fashion. Quantification of G4 and/or  $\gamma$ -H2AX immunostaining in NHAs was blinded.

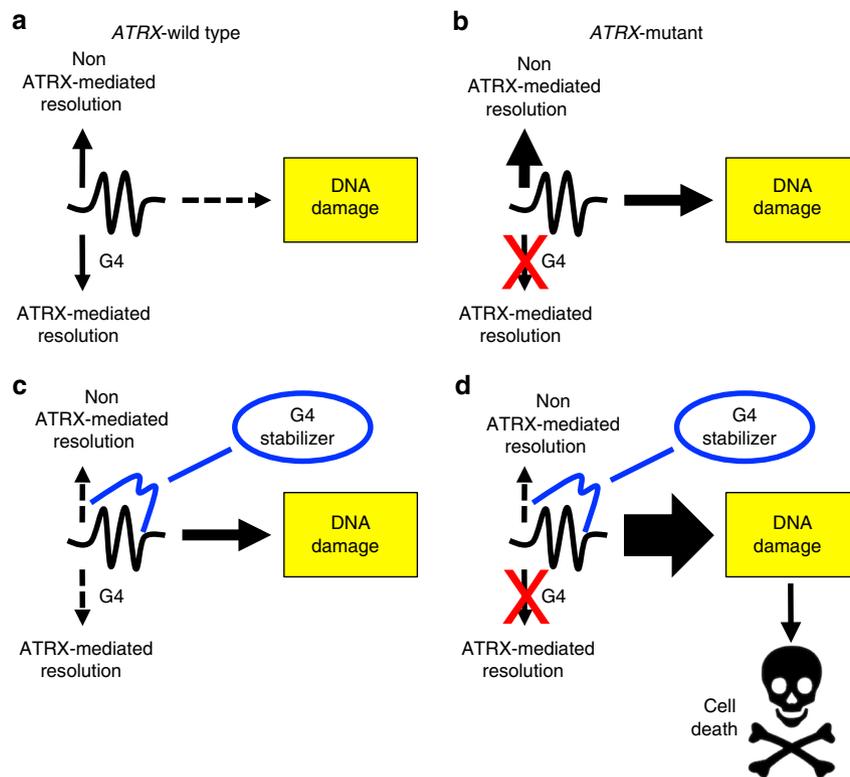
**Antibodies.** All commercially available antibodies used in this study, along with their source and application(s), are listed in Supplementary Table 1.



**Fig. 8** Efficacy of G4 stabilization in vivo is dependent on ATRX deficiency. **a** TS 543 (ATRX intact) xenografts exhibited similar rates of growth when treated with either vehicle control (Veh,  $n = 3$ ) or 12.5 mg/kg CX-3543 ( $n = 3$ ) as reflected by tumor volume over time. **b** Kaplan-Meier analysis of survival in a separate cohort of 10 mice randomized (5 and 5) into two groups, receiving either 12.5 mg/kg CX-3543 or vehicle control (Veh). **c** ATRX knockdown in TS 543 cells (sh590) restored the sensitivity of xenografts to CX-3543 treatment as shown by tumor volume over time in 10 mice randomized (5 and 5) into two groups, receiving either 12.5 mg/kg CX-3543 or vehicle control (Veh). **d** Kaplan-Meier analysis of survival in a separate cohort of 10 mice randomized (5 and 5) into two groups, receiving either 12.5 mg/kg CX-3543 or vehicle control (Veh). **e, f** Representative H&E-stained, immunostained (Ki67 and  $\gamma$ -H2AX), and TEL-FISH stained sections of xenografts from vehicle (Veh) and CX-3543-treated mice harboring either TS 543-shCon (**e**) or TS 543-sh590 (**f**) xenografts. CX-3543-associated histopathological effects were limited to TS 543-sh590 xenografts. Evidence of ALT was not seen in either GSC line.  $P$  values were determined by two-way ANOVA for tumor growth and Log-rank (Mantel-Cox) test for survival curve comparison; scale bars represent 200  $\mu$ m in staining panels and 25  $\mu$ m in Tel-FISH panels

**Cell culture and generation of ATRX-deficient cell lines.** All cell lines used in this study were tested for mycoplasma contamination every three months at the MSKCC Antibody and Bioresource Core. Parental immortalized normal human astrocytes were a gift from R.O. Peiper (UCSF)<sup>61</sup>. TS 543, TS 603, and 08-0537 are patient-derived GSCs<sup>62-64</sup> maintained in NeuroCult™ NS-A Proliferation media (#05751, Stemcell). 08-0537 was generously provided by Hai Yan (Duke).

ATRX knockdown was achieved by introducing either a modified FUGW vector (a gift from David Baltimore (Addgene plasmid # 14883)) carrying an shRNA expression cassette against *ATRX* (shATRX1) (see Supplementary Table 2 for shRNA sequences), a TRIPZ TET-inducible vector (Dharmacon) containing a distinct shRNA against *ATRX* (shATRX2), or a third shRNA against *ATRX* (sh590) from the TRC shRNA library (Sigma). shATRX1- and shATRX2-positive



**Fig. 9** Selectively targeting G4s in the ATRX-deficient context. G4s are normally resolved by both ATRX-dependent and ATRX-independent mechanisms to mitigate DNA damage in cells (**a**). In the setting of ATRX deficiency, DNA damage increases but its cytotoxic effects are dampened by ATRX-independent G4 resolution, maintaining cellular viability (**b**). However, concurrent G4 stabilization impairs these salvage pathways (**c**, **d**) further enhancing DNA damage and inducing cell death in the ATRX-deficient context

cells were FACS-sorted every two passages by fluorescent marker (RFP) for the top 5% of total population to ensure high shRNA expression. Sh590-positive TS 543 cells were subjected to puromycin-based selection.

**Proliferation, cell cycle, and apoptosis analyses.** Flow-cytometry analyses of proliferation and cell cycle were performed using the BD Pharmingen BrdU Flow Kit (# 559619). Apoptosis assays were performed using the Dead Cell Apoptosis Kit (# V13241, Thermo Fisher) with Propidium Iodide (PI) substituted by DAPI to avoid RFP interference.

**In situ visualization of G-quadruplexes,  $\gamma$ -H2AX, 53BP1, and BLM.** The 1H6 antibody was a gift from Dr. Peter M. Lansdorff<sup>65</sup>. The BG4 antibody was purchased from Millipore (MABE917). For immunostaining, cells were grown in chamber slides (Nunc Lab-Tek II, cat no. 154526, Thermo Fisher) and synchronized to G0 phase by 24-h serum starvation. The cells were digested with 10 mg/ml proteinase K for 1 h at 37 °C, followed by fixation (4% paraformaldehyde in PBS for 10 min) and permeabilization (0.5% Tween-20, 0.2% Triton X-100 in PBS, 10 min). To eliminate RNA-structures, cells were treated with 20  $\mu$ g/500  $\mu$ l RNase A (Invitrogen). To confirm specificity towards DNA-G4, cells were incubated in 40 mM Tris Cl (pH 8), 5 mM CaCl<sub>2</sub>, 2 mM MgCl<sub>2</sub>, 100  $\mu$ g/ml BSA alone or including 0.06 U/ $\mu$ l of DNase I (Promega) and 80 gel units/ $\mu$ l of micrococcal nuclease (#M0247S, New England Biolabs) at 37 °C for 2 h. For staining, cells were blocked with goat serum (Sigma) for 4 h at room temperature, then incubated with 1H6 (0.5  $\mu$ g/ml) or BG4 (1:100) at 4 °C overnight. Slides were then washed five times with PBST, incubated with Alexa Fluor 488 or 568 goat anti-mouse IgG (Invitrogen) at room temperature for 2 h, washed five times with PBST and mounted with coverslips using ProLong Gold antifade reagent and DAPI counterstain (Invitrogen). For  $\gamma$ -H2AX monostaining or 53BP1/1H6 or  $\gamma$ -H2AX/BG4 double staining, cells were treated with or without CX-3543 (100 nM), CX-5461 (50 nM), or PDS (2  $\mu$ M) for 3 days prior to synchronization to G0 and incubation with the appropriate primary antibody combinations at 4 °C overnight ( $\gamma$ -H2AX antibody (1:500): # 05-636, Millipore; 53BP1 antibody (1:1000): cat# NB100-304, Novus Biologicals). Secondary antibodies included goat anti-mouse or goat anti-rabbit Alexa Fluor 488 or 568 (1:2000), as appropriate, and were applied as described above. For BLM/ $\gamma$ -H2AX double staining, fixation/permeabilization was performed in ice-cold 100% methanol, and staining conducted sequentially with  $\gamma$ -H2AX antibody (1:500, overnight) and Alexa Fluor 568 goat anti-mouse

antibody (1:2000, 2 h), followed by BLM antibody (Bethyl, A300-110A, 1:50) (overnight) and Alexa Fluor 488 goat anti-rabbit antibody (1:2000, 2 h).

**G4 pulldowns.** Plasmid expressing hf2 was a kind gift of Dr. Shankar Balasubramanian<sup>66,67</sup>. Hf2 antibodies were expressed in BL21 competent cells subjected to 1 mM IPTG induction. The culture supernatant was then bound to Protein A Sepharose (#P9424, Sigma) by chromatography, followed by washing with 50 mM potassium phosphate, 100 mM potassium chloride buffer (pH 7.4) three times. Antibody was then eluted with 0.1 M Tricine buffer (pH 3.0) into 0.1 M potassium phosphate buffer (pH 8.0). For G4 pulldowns, 2  $\mu$ g of hf2 and 50  $\mu$ l of Protein A Dynabeads (#10001D, Thermo Fisher) were mixed and incubated overnight rotating at 4 °C. Beads were washed with PBS five times. Ten  $\mu$ g of genomic DNA from NHAs was sonicated and incubated with beads in 0.5% BSA overnight rotating at 4 °C, followed by six washes with 10 mM Tris pH 7.4, 100 mM KCl, 0.1% Tween-20 and one wash with 10 mM Tris pH 7.4, 100 mM KCl. Bound DNA was eluted in 50  $\mu$ l of 1% SDS, 0.1 M NaHCO<sub>3</sub> at 30 °C for 1 h then purified by QIAquick PCR Purification Kit (Qiagen) to a final volume of 20  $\mu$ l. The recovered DNA was used to determine enrichment of telomeric sequence (Tel1, 2 and X) and the promoter regions of *MYC* and *ZNF618*, using the *ESR1* promoter as a negative control (See Supplementary Table 3 for primer sequences).

**Chromatin immunoprecipitation (ChIP).** NHAs (shCon and shATRX1,  $3 \times 10^7$  cells) were cross-linked with formaldehyde (0.75% v/v, RT, 15 min), quenched with 125 mM glycine (RT, 5 min), harvested, and sonicated with a Bioruptor (Diagenode) in 50 mM HEPES-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8, with 0.1% SDS and proteinase inhibitors, to obtain DNA fragments of 200–300 bp<sup>68</sup>. Fragmented DNA was then subjected to immunoprecipitation with  $\gamma$ -H2AX (2  $\mu$ g) and BLM (5  $\mu$ g) antibodies (see above) and IgG controls at 4 °C overnight with constant agitation. DNA-antibody complexes were then incubated with 20  $\mu$ l Magna Protein G Magnetic Beads (#16-662, Millipore) overnight at 4 °C with constant agitation. Recovered DNA fragments were measured for enrichment at *MYC*, *ZNF618*, and *ESR1* promoters as described above. Additionally, shATRX1 NHAs were treated with PDS (2  $\mu$ M) for 3 days before fixation and ChIP, and enrichment at *MYC*, *ZNF*, and *ESR1* loci was compared to that seen in vehicle treated shATRX1 NHAs.

**TEL-FISH and metaphase cytogenetic analysis.** For cell lines, resuspended cells were incubated with Colcemid (0.1 µg/ml) at 37 °C for 45 min, resuspended in 0.075 M KCl and incubated at 37 °C for 10 min, followed by fixation in methanol:acetic acid (3:1) solution. TEL-FISH was performed according to standard procedures using a CY3-conjugated, telomere-specific nucleic acid probe: 5'-TTAGGGT TAGGGTTAGGG-3' (Applied Biosystems). For xenograft tissues, tumors were removed and subjected to OCT embedding followed by 5 µm sectioning. Frozen sections were fixed with 4% paraformaldehyde for 10 min. After denaturing at 85 °C for 5 min in 10 mM Tris-HCl pH 7.2, 70% formamide, 0.5% blocking solution reagent (Roche), hybridization was performed as described above.

**Cell viability and clonogenic assay.** For standard viability assays, cells (500/well) were incubated with a serial concentration of CX-3543 (10–300 nM), CX-5461 (2.5–500 nM), or PDS (0.1–20 µM) for 7 days in 96-well plates. Cell viability was then assessed with the CellTiter-Glo Luminescent Assay (Promega) according to manufacturer-recommended procedures. To determine clonogenic ability, NHA or TS 543 cells were seeded at 5000 cells/10-cm dish and incubated with vehicle or 50 nM CX-3543 for 14 days. Cells were fixed with 4% paraformaldehyde and stained with 0.005% crystal violet in PBS, followed by three washes in PBS and two washes in ddH<sub>2</sub>O. For soft agar colony formation assays, 50,000 cells were seeded in 6-well plates containing 1% bottom layer and 0.5% top layer soft agar. Cells were then cultured in growth media with or without 50 nM CX-3543. Radiation dosing of 0, 1, 2, or 4 Gy was immediately applied after plating. The 1.5 ml growth media covering the agar cultures was replenished every week. At day 21, colonies were fixed with 4% paraformaldehyde for 30 min and stained with 0.005% crystal violet in PBS overnight. Stained colonies were then washed extensively in PBS and water, and quantified on a Gelcount colony counter (Oxford Optronix).

**SNP arrays.** Genomic DNA was isolated from ATRX-deficient NHAs at passages 5 and 15. As controls, genomic DNA from ATRX-intact parental NHAs was derived at the start point (P0), P5, and P15. Extracted DNA was subjected to Affymetrix Genome-Wide Human SNP 6.0 array analysis (cat# 901182, Thermo Fisher) according to the manufacturer's protocol. Preliminary copy number derivation was facilitated by circular binary segmentation<sup>69,70</sup> to generate CNV segment files with the following information: chromosome, start position, end position, probe number, and segment mean value. For analysis, we focused variations with absolute segment mean value >0.5 for LGG samples and >0.1 for NHA lines. All variations associated with ChrX and ChrY were excluded. CNV length was calculated by using the end position minus the start position. Data were visualized using IGV and GISTIC2.0.

**Xenograft experiments.** All animal protocols and procedures were performed in the xenograft suite at Memorial Sloan-Kettering Cancer Center (Animal protocol # 07-09-015) in accordance with the ethical and experimental regulations of the Institutional Animal Care and Use Committee (IACUC). JHH-273 samples were kind gifts from Dr. Gregory Riggins at the Johns Hopkins University. Tumor samples were mechanically dissociated and small pieces (0.2 mm<sup>3</sup>) were embedded into the flanks of nude mice (Taconic Farms). In parallel, ATRX-intact and ATRX-deficient TS 543 cells at exponential growth phase were dissociated with Accutase (#07920, Stemcell), resuspended in Neurocult media, mixed with Matrigel (#356234, Corning) (1:1) and injected into nude mice flanks in a 50 µl mixture containing 5 × 10<sup>6</sup> cells. Mice were randomized to vehicle or CX-3543 (12.5 mg/kg) treatment groups. Drug delivery occurred via intravenous injection once per day, on a 5 day/week schedule until health-related defined end points. Tumor volumes were measured by calipers and calculated using the formula (l × w<sup>2</sup>)/2, where w is width and l is length in mm. For survival experiments, mice were treated until they reached health-related end points (2000 mm<sup>3</sup> tumor volume). For growth curve comparisons, all mice in a study cohort were sacrificed when the first mouse reached the 2000 mm<sup>3</sup> tumor size threshold. An independent cohort was used for Kaplan–Meier analysis. Xenografted tissues were removed, weighted, and split into two parts. One part was snap frozen for TEL-FISH, while the other half was subjected to FFPE processing. Five-micrometer FFPE sections were deparaffinized and subjected to antigen retrieval. Sections were blocked for non-specific binding with goat serum for 2 h, followed by staining with Ki67 (5 µg/ml, ab15580, Abcam) or γ-H2AX (1:1000, # 05-636, Millipore) antibodies at 4 °C overnight. Sections were washed and incubated with secondary antibody. Ki67 staining were counterstained with Hematoxylin, and γ-H2AX staining were counterstained with DAPI.

**Statistics.** Unless otherwise stated, all results, representing at least three independent experiments, were plotted as mean ± SEM. In general, data were statistically analyzed using unpaired, two-tailed *t*-tests. Log-rank (Mantel–Cox) test were used to determine the significance of differences in Kaplan–Meier analysis of LGG patients and of hind flank xenograft experiments. Two-way ANOVA was used to compare the growth curves of xenografts and the colony formation assays. *P* values are represented using \* for *P* < 0.05, \*\* for *P* < 0.01, \*\*\* for *P* < 0.001, and \*\*\*\* for *P* < 0.0001.

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

**Code availability.** No customized code was used for data processing

### Data availability

All data (raw and processed) and materials related to this manuscript will be made available upon request, utilizing material transfer agreements when appropriate. Raw SNP array data and copy number variation profiles have been deposited in Gene Expression Omnibus (GSE125296), [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE125296>]. Raw western blot data are presented in Supplementary Fig. 12.

Received: 6 June 2018 Accepted: 8 February 2019

Published online: 26 February 2019

### References

- Omuro, A. & DeAngelis, L. M. Glioblastoma and other malignant gliomas: a clinical review. *JAMA* **310**, 1842–1850 (2013).
- Brat, D. J. et al. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N. Engl. J. Med.* **372**, 2481–2498 (2015).
- Jiao, Y. et al. Frequent ATRX, CIC, FUBP1 and IDH1 mutations refine the classification of malignant gliomas. *Oncotarget* **3**, 709–722 (2012).
- Schwartzentruber, J. et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* **482**, 226–231 (2012).
- Ji, J. et al. Inherited germline ATRX mutation in two brothers with ATR-X syndrome and osteosarcoma. *Am. J. Med. Genet. A* **173**, 1390–1395 (2017).
- Masliah-Planchon, J. et al. Does ATRX germline variation predispose to osteosarcoma? Three additional cases of osteosarcoma in two ATR-X syndrome patients. *Eur. J. Hum. Genet.* **26**, 1217–1221 (2018).
- Smolle, M. A. et al. A novel mutation in ATRX associated with intellectual disability, syndromic features, and osteosarcoma. *Pediatr. Blood Cancer* **64**, e26522 (2017).
- Gibbons, R. J. & Higgs, D. R. Molecular-clinical spectrum of the ATR-X syndrome. *Am. J. Med. Genet.* **97**, 204–212 (2000).
- Drane, 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).
- 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. USA* **107**, 14075–14080 (2010).
- 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).
- 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).
- Ratnakumar, K. et al. ATRX-mediated chromatin association of histone variant macroH2A1 regulates alpha-globin expression. *Genes Dev.* **26**, 433–438 (2012).
- Sarma, K. et al. ATRX directs binding of PRC2 to Xist RNA and polycomb targets. *Cell* **159**, 869–883 (2014).
- Voon, H. P. et al. ATRX plays a key role in maintaining silencing at interstitial heterochromatic loci and imprinted genes. *Cell Rep.* **11**, 405–418 (2015).
- Clynes, D. et al. Suppression of the alternative lengthening of telomere pathway by the chromatin remodelling factor ATRX. *Nat. Commun.* **6**, 7538 (2015).
- Heaphy, C. M. et al. Altered telomeres in tumors with ATRX and DAXX mutations. *Science* **333**, 425 (2011).
- 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).
- Clynes, D. et al. ATRX dysfunction induces replication defects in primary mouse cells. *PLoS ONE* **9**, e92915 (2014).
- Koschmann, C. et al. ATRX loss promotes tumor growth and impairs nonhomologous end joining DNA repair in glioma. *Sci. Transl. Med.* **8**, 328ra28 (2016).
- Leung, J. W. et al. Alpha thalassemia/mental retardation syndrome X-linked gene product ATRX is required for proper replication restart and cellular resistance to replication stress. *J. Biol. Chem.* **288**, 6342–6350 (2013).
- 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).

23. Ritchie, K. et al. Loss of ATRX leads to chromosome cohesion and congression defects. *J. Cell Biol.* **180**, 315–324 (2008).
24. Udugama, M. et al. Ribosomal DNA copy loss and repeat instability in ATRX-mutated cancers. *Proc. Natl Acad. Sci. USA* **115**, 4737–4742 (2018).
25. 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).
26. Clynes, D., Higgs, D. R. & Gibbons, R. J. The chromatin remodeller ATRX: a repeat offender in human disease. *Trends Biochem. Sci.* **38**, 461–466 (2013).
27. Paeschke, K., Capra, J. A. & Zakian, V. A. DNA replication through G-quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. *Cell* **145**, 678–691 (2011).
28. Rodriguez, R. et al. Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.* **8**, 301–310 (2012).
29. Watson, L. A. et al. Atrx deficiency induces telomere dysfunction, endocrine defects, and reduced life span. *J. Clin. Invest.* **123**, 2049–2063 (2013).
30. Turcan, S. et al. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* **483**, 479–483 (2012).
31. Turcan, S. et al. Mutant-IDH1-dependent chromatin state reprogramming, reversibility, and persistence. *Nat. Genet.* **50**, 62–72 (2018).
32. Ohba, S. et al. Mutant IDH1 expression drives TERT promoter reactivation as part of the cellular transformation process. *Cancer Res.* **76**, 6680–6689 (2016).
33. Johannessen, T. A. et al. Rapid conversion of mutant IDH1 from driver to passenger in a model of human gliomagenesis. *Mol. Cancer Res.* **14**, 976–983 (2016).
34. Ohba, S., Mukherjee, J., See, W. L. & Pieper, R. O. Mutant IDH1-driven cellular transformation increases RAD51-mediated homologous recombination and temozolomide resistance. *Cancer Res.* **74**, 4836–4844 (2014).
35. Cawthon, R. M. Telomere measurement by quantitative PCR. *Nucleic Acids Res.* **30**, e47 (2002).
36. Hansel-Hertsch, R. et al. G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **48**, 1267–1272 (2016).
37. Pan, Y. F. et al. Regulation of estrogen receptor-mediated long range transcription via evolutionarily conserved distal response elements. *J. Biol. Chem.* **283**, 32977–32988 (2008).
38. Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **43**, 8627–8637 (2015).
39. Sun, H., Karow, J. K., Hickson, I. D. & Maizels, N. The Bloom's syndrome helicase unwinds G4 DNA. *J. Biol. Chem.* **273**, 27587–27592 (1998).
40. Drygin, D. et al. Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. *Cancer Res.* **69**, 7653–7661 (2009).
41. Balasubramanian, S., Hurley, L. H. & Neidle, S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug. Discov.* **10**, 261–275 (2011).
42. Xu, H. et al. CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nat. Commun.* **8**, 14432 (2017).
43. Borodovsky, A. et al. A model of a patient-derived IDH1 mutant anaplastic astrocytoma with alternative lengthening of telomeres. *J. Neurooncol.* **121**, 479–487 (2015).
44. Danussi, C. et al. Atrx inactivation drives disease-defining phenotypes in glioma cells of origin through global epigenomic remodeling. *Nat. Commun.* **9**, 1057 (2018).
45. Rizzo, A. et al. Stabilization of quadruplex DNA perturbs telomere replication leading to the activation of an ATR-dependent ATM signaling pathway. *Nucleic Acids Res.* **37**, 5353–5364 (2009).
46. Berube, N. G. et al. The chromatin-remodeling protein ATRX is critical for neuronal survival during corticogenesis. *J. Clin. Invest.* **115**, 258–267 (2005).
47. Huh, M. S. et al. Stalled replication forks within heterochromatin require ATRX for protection. *Cell Death Dis.* **7**, e2220 (2016).
48. Branzei, D. & Foiani, M. The checkpoint response to replication stress. *Dna Repair. (Amst.)* **8**, 1038–1046 (2009).
49. Petermann, E. & Helleday, T. Pathways of mammalian replication fork restart. *Nat. Rev. Mol. Cell Biol.* **11**, 683–687 (2010).
50. Pflumm, M. F. The role of DNA replication in chromosome condensation. *Bioessays* **24**, 411–418 (2002).
51. De, S. & Michor, F. DNA secondary structures and epigenetic determinants of cancer genome evolution. *Nat. Struct. Mol. Biol.* **18**, 950–955 (2011).
52. von Deimling, A. et al. Diffuse astrocytoma, IDH-mutant. in *WHO classification of tumours of the central nervous system* (eds Louis, D. N., Ohgaki, H., Wiestler, O. D. & Cavenee, W. K.) (International agency for research on cancer (IARC), Lyon, 2016).
53. Helleday, T. The underlying mechanism for the PARP and BRCA synthetic lethality: clearing up the misunderstandings. *Mol. Oncol.* **5**, 387–393 (2011).
54. Laperriere, N., Zuraw, L. & Cairncross, G. Radiotherapy for newly diagnosed malignant glioma in adults: a systematic review. *Radiother. Oncol.* **64**, 259–273 (2002).
55. Walker, M. D. et al. Evaluation of BCNU and/or radiotherapy in the treatment of anaplastic gliomas. A cooperative clinical trial. *J. Neurosurg.* **49**, 333–343 (1978).
56. Walker, M. D. et al. Randomized comparisons of radiotherapy and nitrosoureas for the treatment of malignant glioma after surgery. *N. Engl. J. Med.* **303**, 1323–1329 (1980).
57. Wiestler, B. et al. ATRX loss refines the classification of anaplastic gliomas and identifies a subgroup of IDH mutant astrocytic tumors with better prognosis. *Acta Neuropathol.* **126**, 443–451 (2013).
58. Cairncross, J. G. et al. Benefit from procarbazine, lomustine, and vincristine in oligodendroglial tumors is associated with mutation of IDH. *J. Clin. Oncol.* **32**, 783–790 (2014).
59. Zheng, X. H. et al. A cisplatin derivative tetra-Pt(bpy) as an oncotherapeutic agent for targeting ALT cancer. *J. Natl Cancer Inst.* **109**, dxj061 (2017).
60. Guilbaud, G. et al. Local epigenetic reprogramming induced by G-quadruplex ligands. *Nat. Chem.* **9**, 1110–1117 (2017).
61. Sonoda, Y. et al. Formation of intracranial tumors by genetically modified human astrocytes defines four pathways critical in the development of human anaplastic astrocytoma. *Cancer Res.* **61**, 4956–4960 (2001).
62. Turcan, S. et al. Efficient induction of differentiation and growth inhibition in IDH1 mutant glioma cells by the DNMT Inhibitor Decitabine. *Oncotarget* **4**, 1729–1736 (2013).
63. Waitkus, M. S. et al. Adaptive evolution of the GDH2 allosteric domain promotes gliomagenesis by resolving IDH1(R132H)-induced metabolic liabilities. *Cancer Res.* **78**, 36–50 (2018).
64. Silber, J. et al. miR-34a repression in proneural malignant gliomas upregulates expression of its target PDGFRA and promotes tumorigenesis. *PLoS ONE* **7**, e33844 (2012).
65. Henderson, A. et al. Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.* **42**, 860–869 (2014).
66. Lam, E. Y., Beraldi, D., Tannahill, D. & Balasubramanian, S. G-quadruplex structures are stable and detectable in human genomic DNA. *Nat. Commun.* **4**, 1796 (2013).
67. Fernando, H., Rodriguez, R. & Balasubramanian, S. Selective recognition of a DNA G-quadruplex by an engineered antibody. *Biochemistry* **47**, 9365–9371 (2008).
68. Komata, M. et al. Chromatin immunoprecipitation protocol for mammalian cells. *Methods Mol. Biol.* **1164**, 33–38 (2014).
69. Korn, J. M. et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat. Genet.* **40**, 1253–1260 (2008).
70. Venkatraman, E. S. & Olshen, A. B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**, 657–663 (2007).

## Acknowledgements

We would like to thank Drs. Peter Lansdorp and Shankar Balasubramanian for providing the 1H6 and hf2 antibodies, respectively. We would also like to acknowledge Dr. Cameron Brennan for providing the TS 543 GSC line. Finally, we would like to acknowledge the Genomics Core Facility at the Albert Einstein College of Medicine for their assistance in performing SNP arrays, and the Molecular Cytogenetics Core at MSKCC for Tel-FISH. J.T.H. is supported by a Research Scholars Grant, RSG-16-179-01-DMC, from the American Cancer Society. Support for this work also came from the Sontag Foundation (J.T.H.), the Sidney Kimmel Foundation (J.T.H.), and Cycle for Survival (J.T.H.). We acknowledge support from two NIH/NCI Cancer Center Support Grants (CCSGs) for MDACC (P30 CA016672) and MSKCC (P30 CA08748).

## Author contributions

Conceptualization and design were done by Y.W., T.A.C. and J.T.H. Development and methodology were done by Y.W., T.A.C. and J.T.H. Acquisition of data was done by Y.W., A.T.W., W.H.W., R.S. and C.D. Analysis and interpretation of data were done by Y.W., J.Y., K.K., T.A.C. and J.T.H. Writing, review, and/or revision of the manuscript were done by Y.W., K.K., T.A.C. and J.T.H. Administrative, technical, or material support were done by G.J.R., K.K., E.P.S., T.A.C. and J.T.H. Study supervision was done by T.A.C. and J.T.H.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-019-08905-8>.

**Competing interests:** T.A.C. is a co-founder of Gritstone Oncology and holds equity. He holds equity in An2H. He acknowledges grant funding from Bristol-Myers Squibb, AstraZeneca, Illumina, Pfizer, An2H, and Eisai, and he has served as a paid advisor for Bristol-Myers Squibb, Illumina, Eisai, and An2H. M.S.K. has licensed the use of TMB for the identification of patients that benefit from immune checkpoint therapy to PGDx and T.A.C. receives royalties as part of this licensing agreement. The remaining authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Journal peer review information:** *Nature Communications* thanks Maria Castro and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

# Stalled replication forks within heterochromatin require ATRX for protection

MS Huh<sup>1,4</sup>, D Ivanochko<sup>1,2,4</sup>, LE Hashem<sup>1,3</sup>, M Curtin<sup>1,2</sup>, M Delorme<sup>1,2</sup>, E Goodall<sup>1,2</sup>, K Yan<sup>1</sup> and DJ Picketts<sup>\*1,2,3</sup>

Expansive growth of neural progenitor cells (NPCs) is a prerequisite to the temporal waves of neuronal differentiation that generate the six-layered neocortex, while also placing a heavy burden on proteins that regulate chromatin packaging and genome integrity. This problem is further reflected by the growing number of developmental disorders caused by mutations in chromatin regulators. ATRX gene mutations cause a severe intellectual disability disorder ( $\alpha$ -thalassaemia mental retardation X-linked (ATRAX) syndrome; OMIM no. 301040), characterized by microcephaly, urogenital abnormalities and  $\alpha$ -thalassaemia. Although the ATRX protein is required for the maintenance of repetitive DNA within heterochromatin, how this translates to disease pathogenesis remain poorly understood and was a focus of this study. We demonstrate that *Atrx*<sup>FoxG1Cre</sup> forebrain-specific conditional knockout mice display poly(ADP-ribose) polymerase-1 (Parp-1) hyperactivation during neurogenesis and generate fewer late-born Cux1- and Brn2-positive neurons that accounts for the reduced cortical size. Moreover, DNA damage, induced Parp-1 and Atm activation is elevated in progenitor cells and contributes to their increased level of cell death. ATRX-null HeLa cells are similarly sensitive to hydroxyurea-induced replication stress, accumulate DNA damage and proliferate poorly. Impaired BRCA1-RAD51 colocalization and PARP-1 hyperactivation indicated that stalled replication forks are not efficiently protected. DNA fiber assays confirmed that MRE11 degradation of stalled replication forks was rampant in the absence of ATRX or DAXX. Indeed, fork degradation in ATRX-null cells could be attenuated by treatment with the MRE11 inhibitor mirin, or exacerbated by inhibiting PARP-1 activity. Taken together, these results suggest that ATRX is required to limit replication stress during cellular proliferation, whereas upregulation of PARP-1 activity functions as a compensatory mechanism to protect stalled forks, limiting genomic damage, and facilitating late-born neuron production.

*Cell Death and Disease* (2016) 7, e2220; doi:10.1038/cddis.2016.121; published online 12 May 2016

Mutations in genes encoding epigenetic regulators are the cause of many neurodevelopmental disorders, thereby highlighting the importance of chromatin remodeling to progenitor cell growth, competency, cell fate, and differentiation capacity.<sup>1</sup> In this regard, mutations in the human ATRX gene cause  $\alpha$ -thalassaemia mental retardation X-linked (ATRAX; OMIM no. 301040) syndrome, a severe intellectual disability disorder commonly associated with urogenital abnormalities, facial dysmorphism, and  $\alpha$ -thalassaemia.<sup>2,3</sup>

The ATRX gene encodes a 280 kDa protein with two chromatin-interaction domains, a C-terminal SNF2 helicase-like domain that provides DNA-dependent ATPase activity and an N-terminal ADD (ATRAX-DNMT3-DNMT3L) domain that serves as a dual histone modification recognition module (H3K9me3/H3K4me0; H3K9me3/H3S10p) to target ATRX to heterochromatin.<sup>4–6</sup> Moreover, ATRX interacts with DAXX to form a histone chaperone complex that loads histone H3.3 onto telomeres, imprinted genes, and endogenous retroviral

elements, to establish and maintain a heterochromatin environment.<sup>7–11</sup> Nonetheless, it remains unclear how these biochemical functions contribute to brain development.

Forebrain-specific inactivation of *Atrx* in mice results in enhanced apoptosis and cerebral hypocellularity,<sup>12</sup> a phenotypic feature commonly observed in ATRAX patients.<sup>13</sup> Further characterization of proliferating cells lacking *Atrx* demonstrate that S-phase progression is delayed and accompanied with an activated DNA-damage response, fragile telomeres, and mitotic catastrophe that enhances cell death in rapidly expanding progenitors of the testis, skeletal muscle, and CNS.<sup>12,14–16</sup> Aberrant replication of heterochromatin was suggested by ChIP-Seq analysis as *Atrx* binding sites are enriched at simple repeats, including telomeres and other guanine-rich sequences with a propensity to form G4 quadruplexes.<sup>17</sup> Moreover, it was proposed that disease pathogenesis could arise from an inability to prevent G4-quadruplex formation, which would impede replication and transcription.<sup>18,19</sup> Initial support for this

<sup>1</sup>Regenerative Medicine Program, Ottawa Hospital Research Institute, Ottawa, ON K1H 8L6, Canada; <sup>2</sup>Department of Biochemistry, Microbiology, and Immunology, Faculty of Medicine, University of Ottawa, Ottawa, ON K1H 8M5, Canada and <sup>3</sup>Department of Cellular and Molecular Medicine, Faculty of Medicine, University of Ottawa, Ottawa, ON K1H 8M5, Canada

\*Corresponding author: DJ Picketts, Regenerative Medicine Program, Ottawa Hospital Research Institute, 501 Smyth Road, Ottawa, ON K1H 8L6, Canada. Tel: +1 613 737 8989; Fax: +1 613 737 8803; E-mail: dpicketts@ohri.ca

<sup>4</sup>These authors contributed equally to this work.

**Abbreviations:** ADD, ATRX-DNMT3-DNMT3L; ATRX,  $\alpha$ -thalassaemia mental retardation X-linked; DSB, double strand break; dsDNA, double-stranded DNA; HP1 $\alpha$ , heterochromatin protein 1 $\alpha$ ; HU, hydroxyurea; IF, immunofluorescent; IZ, intermediate zone; KD, knockdown; MRN, MRE11-RAD50-NBS1; NPC, neural progenitor cell; PAR, polyADP-ribose; PARP-1, Poly(ADP-Ribose) polymerase 1; pATM, phosphorylated ataxia telangiectasia-mutated; shRNA, short hairpin ribonucleic acid; siRNA, small interfering RNA; SVZ, subventricular zone; VZ, ventricular zone

Received 10.2.16; revised 04.4.16; accepted 08.4.16; Edited by A Oberst

model came from studies showing that *Atrx* interacts with the Mre11-Rad50-Nbs1 (MRN) complex and that *Atrx*-deficient cells have an increase in stalled replication forks.<sup>15,20</sup> Mechanisms that protect stalled replication forks are especially critical during mid-late S phase, because of the abundance of natural barriers present in heterochromatin.<sup>21</sup>

Here, we examined whether *Atrx* functions to protect stalled replication forks from collapse and subsequent DNA damage. Indeed, we observed that *Atrx*-deficient cells acquire DNA damage in the S phase, which persists and accumulates in a cell-cycle progressive manner. The replication stress is defined by reduced colocalization of BRCA1 with RAD51, indicating aberrant replication fork protection. The degradation of replication forks is mediated by Mre11, which leads to an increase in double-strand DNA (dsDNA) breaks, fork collapse, genomic instability, and cell death that reduces the progenitor cell pool. As a consequence of fork degradation, neural progenitors activate poly(ADP-ribose) polymerase-1 (Parp-1) to promote fork protection and cell survival, thereby limiting upper layer neuron loss. Indeed, PARP-1 inhibition further perturbed cell growth. Moreover, acute knockdown (KD) of Daxx resulted in a similar degradation of nascent DNA strands, suggesting that histone H3.3 loading facilitates replication fork protection.

## Results

**Increased DNA damage in neural progenitors compromises late-born neuron production.** Previous work in our lab demonstrated that *Atrx*-null primary myoblasts were incapable of prolonged expansion owing to the S-phase defects and genomic instability that severely compromised muscle regeneration.<sup>16</sup> If forebrain progenitor expansion was similarly affected, we reasoned that early-born neuron production would not be compromised but later born neuron production would be decreased, resulting in the reduced cortical mass we observed in *Atrx*<sup>FoxG1Cre</sup> forebrain-specific conditional knockout (*Atrx* cKO) mice.<sup>12</sup> To assess neuron production in *Atrx* cKO mice, we determined the proportion of cells comprising the different cortical layers using layer-specific markers. The earliest born neurons comprise the subplate and the deep layers (VI and V) of the cortex as the forebrain is generated in an inside-out manner. We observed a significant proportional increase in Nurr1+ subplate neurons but no differences in the layer VI (Tbr1+), layer V (Ctip2+), or layer IV (Foxp1+) cells in the *Atrx* cKO brains compared with wild-type (WT) littermates (Figure 1a and Supplementary Figure 1). While this suggested that a sufficient progenitor pool existed to generate the early-born neurons, we observed a significant reduction in the latest born Cux1+ neurons (layer II/III), whereas Brn2+ and Satb2+ neurons showed reduced levels that did not reach statistical significance (Figure 1b). Moreover, the cerebral cortex of *Atrx* cKO mice contained significantly fewer neurons than their WT littermates at E18.5 (Figure 1c), indicating that progenitor cell expansion was compromised.

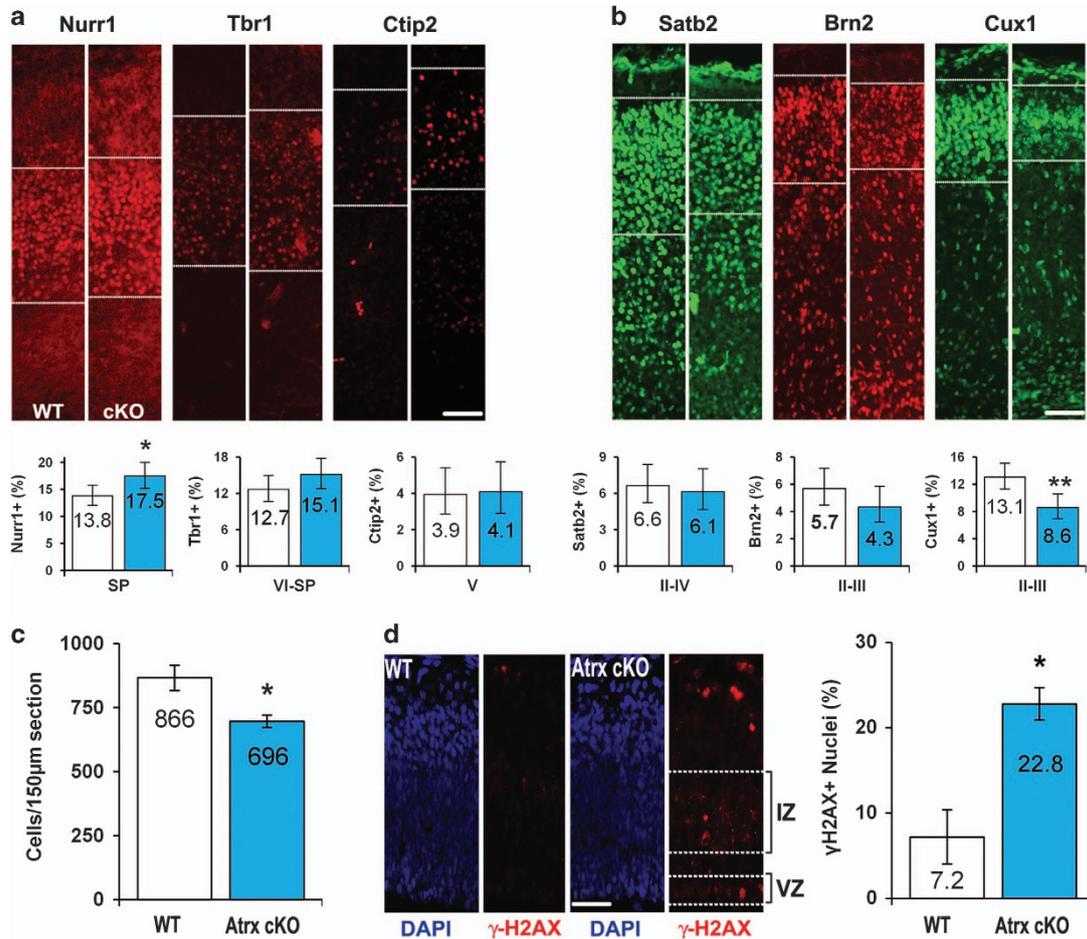
To determine whether genome instability might be the cause of reduced neuron production, we examined the DNA-damage marker  $\gamma$ H2AX by immunofluorescent (IF) staining of E13.5

cortical sections. We observed a significant increase in  $\gamma$ H2AX+ cells that was predominantly located in the proliferative ventricular (VZ) and intermediate (IZ) zones (Figure 1d). Furthermore, we observed an accumulation of genomic damage by E15.5 as assessed by the colocalization of  $\gamma$ H2AX signaling with markers for radial glial (Pax6+) and intermediate (Tbr2+) progenitor cells (Supplementary Figure 2). As the genomic instability in *Atrx* cKO myoblasts was caused by DNA replication stress, we examined Parp-1 activity, a known effector of this pathway. Parp-1 activity was assessed using antibodies specific to Parp-1 and polyADP-ribose (PAR), the moiety added to substrates when the polymerase is active. IF staining of E13.5 *Atrx* cKO neocortices revealed increased PAR staining primarily within the proliferative zone (Figures 2a and b). Immunoblots from cortical extracts demonstrated that this was not due to changes in Parp-1 expression but increased activity (Figure 2c). Indeed, a high level of PARylation was observed at E12.5 and E13.5 in all embryos but it persisted only in the *Atrx* cKO embryos at E14.5 and E15.5 (Figure 2c). As such, we used the E13.5 cortical extracts to assess the activation of the DNA-damage response via phosphorylation of ataxia telangiectasia-mutated (pATM) and H2AX ( $\gamma$ H2AX). Both mutant and WT samples showed active PARylation, but only *Atrx* cKO extracts showed increased pATM and  $\gamma$ H2AX to indicate an activated DNA-damage response (Figure 2c). Interestingly, the Parp-1 immunoblots show a shift in size only in the mutant lanes that probably reflects significant auto-PARylation of the Parp-1 protein (Figure 2c and Supplementary Figure 4). As an indication that DNA damage was leading to cell death, we harvested embryonic cortical extracts from *Atrx* cKO and WT littermates at E12.5 and E17.5 for caspase activity assays. We observed a significant increase in the activation of the executioner caspase, caspase-3, that was mediated by an intrinsic response, as we observed an increase in caspase-9 activity but not caspase-8 (Supplementary Figure 3).

Collectively, these data suggest that genomic instability within the neural precursor population contributes to the observed neuronal cell loss. As depicted in Figure 2d, we postulate that genomic damage accumulates with each successive pass through S phase in the *Atrx*-null progenitor cells, and with seven to eight cell cycles within the span of 3 days there is diminished viability, thereby reducing the pool of late-stage progenitors that generate the upper layer neurons.

## Delayed S phase in ATRX KD cells leads to increased activation of p53-ATM checkpoint in the subsequent G1.

To further investigate the mechanisms by which ATRX regulates genomic stability, we generated both acute and stable KD HeLa cells using siATR<sub>X</sub> or short hairpin-expressing plasmids (psiRNA ATR<sub>X</sub>) with their respective controls (siScrambled (siScram) and psiRNA LacZ). Cell cycle progression analysis of BrdU-labeled cells revealed that psiRNA ATR<sub>X</sub> cells were delayed through S and G2-M phase, similar to primary myoblasts (Supplementary Figure 5; Huh *et al.*<sup>16</sup>). As extended passaging of our psiRNA ATR<sub>X</sub>-stable clones resulted in the selective suppression of the shRNA ATR<sub>X</sub> transgene, the remainder of our experiments used the acute KD model. Following transfection, protein levels of ATR<sub>X</sub> were nearly undetectable by 48 h and

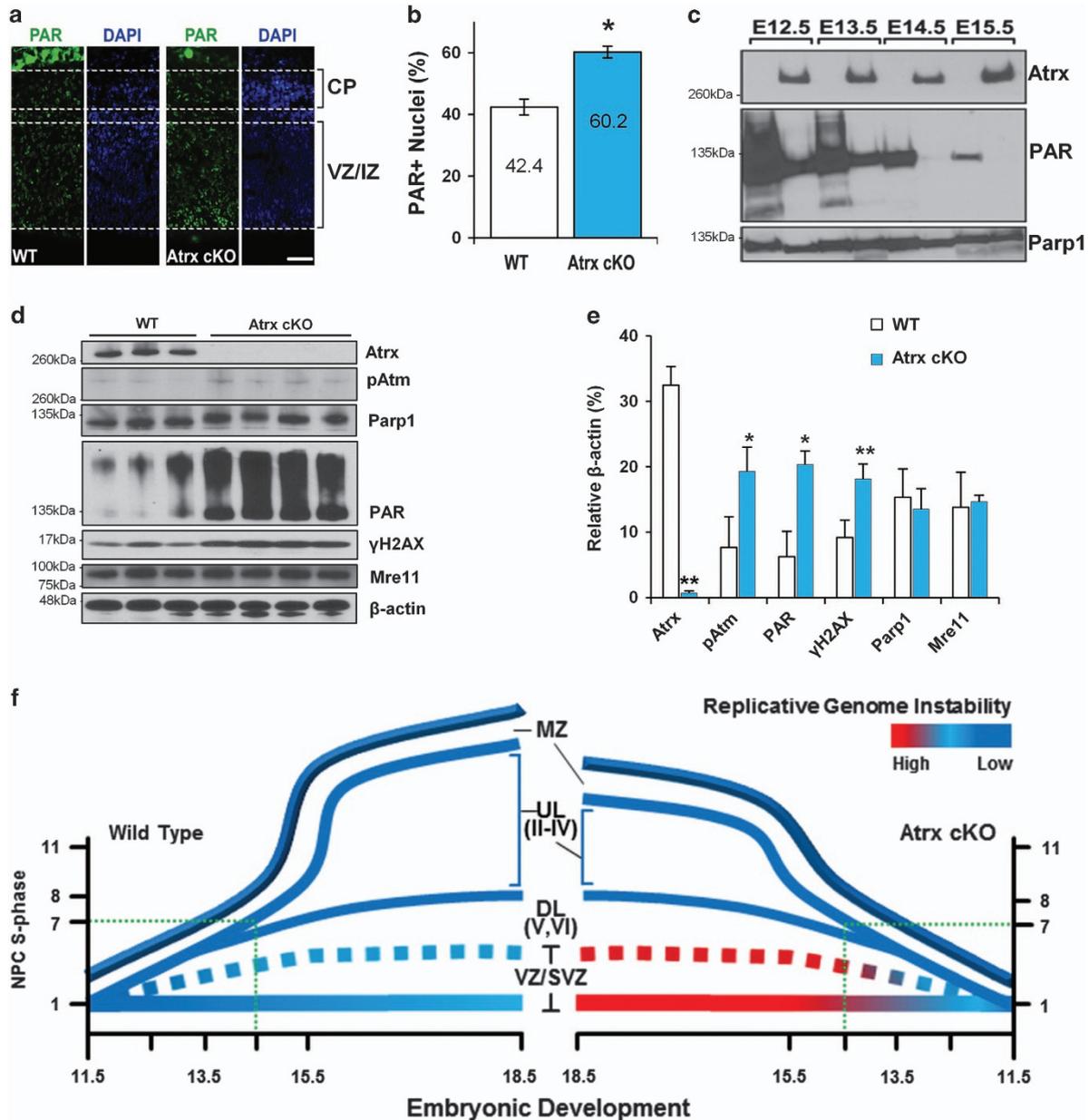


**Figure 1** *Atrx* facilitates the production of late-born cortical neurons by preventing genomic instability in neural precursor cells. Representative micrographs and quantification of neurons located in the deep (a) or upper (b) neocortical layers from E18.5 *Atrx* cKO and WT coronal brain sections. Sections were probed with antibodies that specifically labeled the subplate (SP; *Nurr1*), layer VI-SP (*Tbr1*), and layer V (*Ctip2*), layers II-IV (*Satb2*), and layers II/III (*Brn2* and *Cux1*). Labeled neurons within bounded areas were quantified as a percent of total nuclei within the neocortex. Values represent percent total  $\pm$  95% CI. \* $P < 0.05$  by z-score, whereas \*\* $P < 0.01$  by z-score;  $\times 200$  magnification. Scale bar, 100  $\mu$ m. (c) Average cell density counts from E18.5 WT and *Atrx* cKO cortical sections following DAPI staining. (d) Representative IF micrographs of E13.5 *Atrx* cKO and WT embryos coronal brain sections stained for  $\gamma$ -H2AX (red) or counterstained with DAPI (blue) to label all nuclei. NPCs reside in the VZ and IZ, as indicated by dotted lines;  $\times 200$  magnification. Scale bar, 100  $\mu$ m. Values represent proportional mean  $\pm$  S.E.M. \* $P < 0.05$  by Student's *t*-test

remained absent until 120 h, while we also observed an increase in  $\gamma$ H2AX signaling over this timeline (Supplementary Figure 6). As such, this model is able to replicate our *in vivo* results and can be used to explore the role of ATRX during replication stress.

Previous work has demonstrated that *Atrx*-null cells are delayed through S phase and have an increased incidence of stalled replication forks.<sup>15,16,20</sup> As stalled replication forks often collapse and form dsDNA breaks,<sup>22</sup> we reasoned that the cell loss observed in ATRX KD cells may be due to the progressive accumulation of double strand breaks (DSBs) during progenitor proliferation. For this study, we examined the activation status of ATM with respect to cell-cycle stage (S/G2 or G1) at 72 and 96 h post-transfection. In this regard, cells were costained for pATM and cyclin A (Figure 3a). To quantify pATM signaling pertaining to DNA damage, cells with punctate staining were scored, while cytoplasmic pATM+ cells were excluded, as these represent cells undergoing mitosis.<sup>23,24</sup> Similarly, cells transiting S/G2 phases of the cell cycle were

distinguished by cyclin A staining,<sup>25,26</sup> and this was confirmed in our hands (Supplementary Figure 7). At both the 72 and 96 h time points, we observed a significant increase in the proportion of ATRX KD cells (45.8% and 48.5%, respectively) with focal pATM nuclear staining compared with siScram (39.3% and 36.1%, respectively) control cells (Figure 3b). When total pATM cell counts were dissected into cells in S/G2 (cyclin A+) or G1 (cyclin A-) phase of the cell cycle, we observed a > 50% increase in pATM staining in S/G2 phase at both 72 and 96 h (Figure 3c). Interestingly, we observed a time-dependent increase in pATM staining in the ATRX KD cells within the G1 sub-populations. The ATRX KD and control cells showed no difference at 72 h, but at 96 h post-transfection focal pATM staining significantly increased (compare 38.8% versus 29.8%) in the ATRX KD cells (Figure 3d). These findings illustrate the persistence and accumulation of a replication-dependent DDR response in the subsequent G1 of ATRX KD cells. Moreover, it further supports the model that progenitors accumulate more DSBs, ultimately



**Figure 2** Enhanced activation of DNA-damage response pathways in *Atrx* cKO neuroprogenitors. (a) Representative IF micrographs of E13.5 coronal cortical sections from *Atrx* cKO and WT embryos stained with poly(ADP-ribose) antibodies (PAR; green) and counterstained with DAPI (blue). The cortical plate (CP) and NPC proliferative zones (VZ/IZ) are marked by dotted lines;  $\times 200$  magnification. Scale bar,  $100 \mu\text{m}$ . (b) Quantification of PAR-positive nuclei shown in (a). Values represent the mean  $\pm$  S.E.M.;  $n = 3$ ;  $*P < 0.05$  by Student's *t*-test. (c) Protein extracts from *Atrx* cKO and WT cortices were harvested daily from E12.5 until E15.5 and immunoblotted for Parp-1 activity (PAR), Parp-1 or Atrx. (d) Immunoblot analysis for DNA-damage signaling in E13.5 cortical extracts from WT ( $n = 3$ ) and *Atrx* cKO ( $n = 4$ ) embryos. (e) Densitometry quantification of blot shown in (d). Values are the mean  $\pm$  S.E.M.  $*P < 0.05$ ;  $**P < 0.01$ , by Student's *t*-test. (f) Developmental model of replicative stress induced loss of late-born neurons in the *Atrx* cKO mice. The X axis shows the developmental time and the Y axis shows the number of cycles the NPCs have undergone. Blue lines depict the generation of deep (DL) and upper layer (UL) neurons. Dotted green lines indicate the timing of progenitor cell loss. At this point, progenitors from *Atrx* cKO mice within the VZ/SVZ (red line) have high levels of genomic damage that compromise their survival, resulting in a smaller cortex by E18.5

resulting in genomic instability and activation of cell death pathways.

**Impaired RAD51 colocalization to BRCA1 foci in ATRX KD cells.** Heterochromatin contains an abundance of simple repeats that are prone to instability during replication, forming unusual DNA structures (e.g. cruciform, Z-DNA, and

G-quadruplexes) that can cause replication fork stalling.<sup>21,27</sup> ATRX is a heterochromatin-associated protein that preferentially binds to G-rich tandemly repeated DNA sequences that form G-quadruplexes.<sup>17,28</sup> As such structures require homology-directed recombination (HR) repair to remove them,<sup>29</sup> we hypothesized that the absence of ATRX during replication may compromise the function of the HR

machinery at replicating heterochromatin. In this regard, both ATR<sub>X</sub> and BRCA1 colocalized to replicating heterochromatin domains marked by either heterochromatin protein 1 $\alpha$  (HP1 $\alpha$ ) or mid-late S-phase BrdU-labeled foci (Supplementary Figures 8A and B). To assess whether there was active HR repair after ATR<sub>X</sub> KD, we colabeled cells with BRCA1 and Rad51, functional beacons for HR machinery recruitment at sites of stalled replication forks.<sup>22,30–32</sup> Double IF detection of BRCA1 and RAD51 revealed colocalized nuclear focal signals (Figure 3e). Quantification of BRCA1 foci revealed a greater number of BRCA1 foci present in ATR<sub>X</sub> KD cells *versus* controls (compare 11.5 with 7.4 foci per nucleus respectively; Figure 3f). Despite this overall increase in the frequency of BRCA1 foci, the proportion of BRCA1 foci with colocalized RAD51 signals were markedly reduced in ATR<sub>X</sub> KD cells by 31% relative to controls (Figure 3h). Taken together, these data suggest that insufficient loading of RAD51 at BRCA1 foci may compromise HR-mediated fork restart or stability in the absence of ATR<sub>X</sub>.

**PARP-1 activation functions as a compensatory protective response to stalled replication forks.** We next questioned whether the increased PAR activity we observed in the *Atrx* cKO forebrain indicated a compensatory mechanism to protect stalled replication forks upon RAD51 dysregulation. PARPs are multifunctional enzymes that affect DNA repair, replication fork protection, and restart.<sup>24,33–35</sup> Moreover, PARP-1 hyperactivation in cells with compromised HR pathways has been attributed to a protective response induced by stalled and collapsed replication forks.<sup>36,37</sup> We first confirmed that increased PAR signaling was also detected in ATR<sub>X</sub> KD cells, while total PARP-1 levels remained unchanged (Figure 4a, compare lanes 3 and 1). In addition, we used siPARP-1 to attribute increased PARylation specifically to PARP-1. Indeed, PARP-1 accounts for ~90% of PARylation,<sup>38</sup> and we observed a marked decrease in PAR signaling when cells were treated with both siATR<sub>X</sub> and siPARP-1 (Supplementary Figure 10). As other studies have shown that HR-deficient cells are commonly hypersensitive to PARP-1 inhibition,<sup>39</sup> we used the PARP-1 inhibitor PJ34 to assess whether the ATR<sub>X</sub> KD cells were similarly sensitive. PARP-1 inhibition by PJ34 potently suppressed PAR signaling in ATR<sub>X</sub> KD cells, with a concomitant increase in 53BP1 protein levels compared with siScram controls (Figure 4a, lanes 4 and 3). Quantification of 53BP1-positive nuclei revealed an 83% increase in frequency within PJ34-treated ATR<sub>X</sub> KD cells relative to PJ34-treated controls (Figure 4b and Supplementary Figure 9A). Moreover, PJ34-treated ATR<sub>X</sub> KD cells showed an increased level of TUNEL+ (terminal uridine deoxynucleotidyl transferase dUTP nick-end labeling-positive) nuclei and a severe attenuation of their growth rate over a 5-day time course measured with a WST-1 cell viability assay (Figures 4c–e). Taken together, these experiments suggest that increased PARP-1 activity observed in the absence of ATR<sub>X</sub> represents a protective response to maintain the integrity of stalled replication forks.

**The ATR<sub>X</sub>-DAXX complex facilitates replication fork processivity and protection.** ATR<sub>X</sub>-depleted ES cells exhibit a greater sensitivity to hydroxyurea (HU)-induced

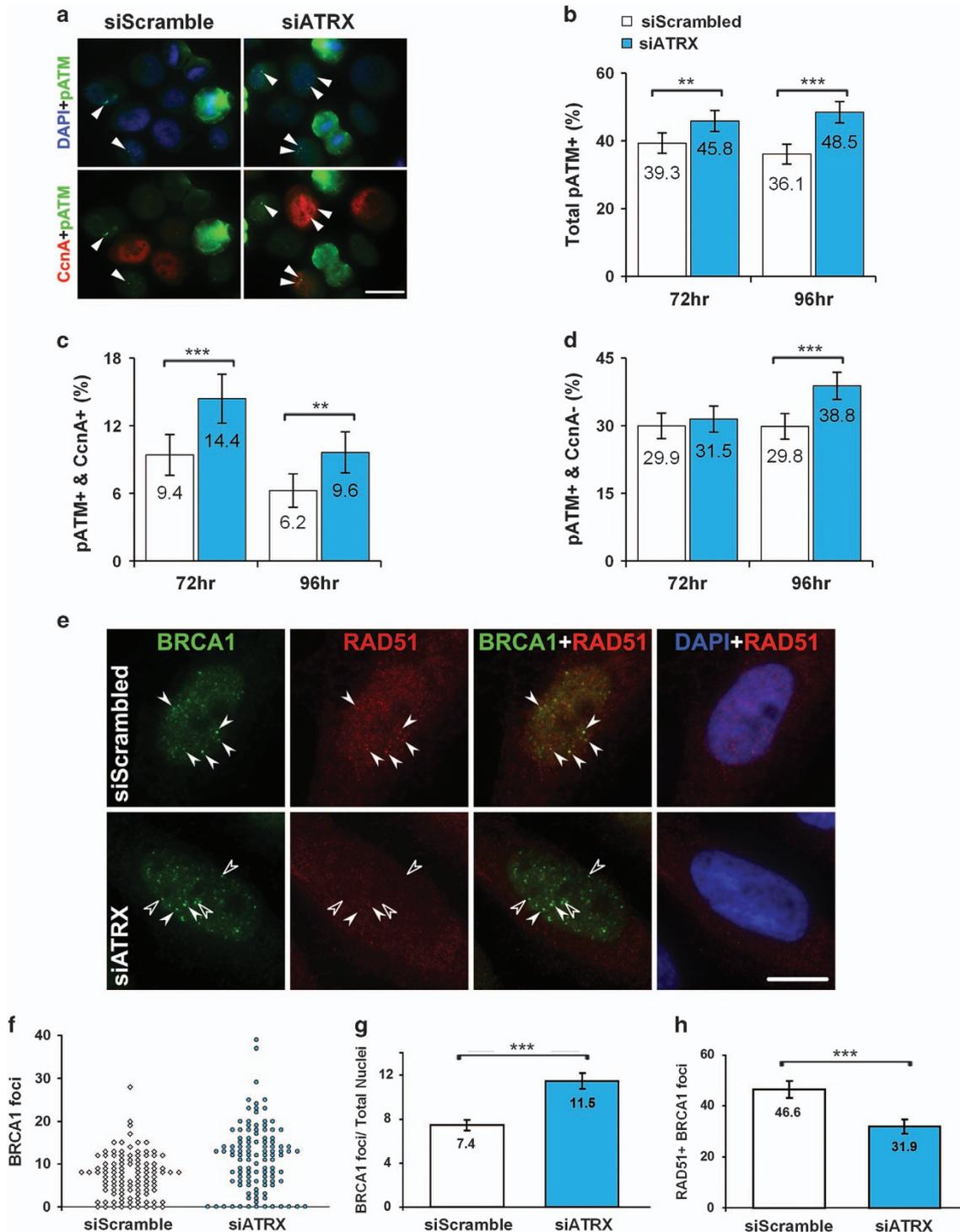
replication fork stalling and delayed replication restart.<sup>15,20</sup> These studies also identified a physical interaction between ATR<sub>X</sub> and the MRN complex.<sup>15,20</sup> However, the mechanism causing the increased fork stalling was not determined. Based on reduced Rad51 colocalization with BRCA1 and active PARP-1, we reasoned that replication fork protection could be compromised. In this regard, HR proteins such as BRCA1/2, RAD51, and MRE11 are functionally critical for the protection of stalled replication forks, independent of their role in dsDNA repair.<sup>40</sup> RAD51 nucleofilament formation at stalled replication forks prevents MRE11-dependent degradation of newly synthesized DNA to allow for the resumption of DNA synthesis.<sup>41</sup> Indeed, artificially blocking RAD51 nucleofilament formation by overexpressing the RAD51 binding peptide BRC4 potently induced fork destabilization upon HU exposure.<sup>30</sup> To assess whether MRE11 exonuclease activity was overly active, we performed DNA fiber studies following HU-induced replication fork stalling, with or without ATR<sub>X</sub> present. Previous work has implicated BRCA1 in the protection of stalled replication forks.<sup>30</sup> Indeed, we confirmed that BrdU-labeled nascent replication tracts of BRCA1-deficient cells (siBRCA1) were markedly shorter following HU treatment compared with controls (Supplementary Figure 11). Quite strikingly, nascent replication tracts in ATR<sub>X</sub> KD cells were equally as short as the tracts observed in BRCA1 KD cells (Supplementary Figure 11). Shorter BrdU-labeled nascent DNA tracts may be the result of decreased replisome processivity rates and/or the instability to protect nascent strands from degradation at sites of stalled forks. To delineate the contribution of these processivity mechanisms, DNA track lengths were compared between the ATR<sub>X</sub> KD cells and siScram control cells without HU-induced fork stalling. While we observed that ATR<sub>X</sub> KD cells produced significantly shorter tracks than siScram control cells, track length reduction was significantly exacerbated upon HU treatment, indicating that fork protection is also compromised (Figure 5a). Moreover, chemical inhibition of MRE11 with the small molecule mirin has been demonstrated to protect stalled replication forks from exonuclease resectioning.<sup>35,42</sup> Indeed, mirin treatment of ATR<sub>X</sub> KD cells produced mean replication tract lengths that were comparable to that of controls (Figure 5b), suggesting that ATR<sub>X</sub> mediates MRE11-dependent degradation at stalled replication forks. Accordingly, ATR<sub>X</sub> may directly suppress MRE11-dependent degradation at stalled forks as it co-immunoprecipitates with both MRE11 and NBS1 in WT asynchronous cells (Supplementary Figure 12A). Regardless, H3.3 has been shown to facilitate replication fork processivity during replication stress and the ATR<sub>X</sub>-DAXX complex serves as a chaperone for loading this histone variant.<sup>43,44</sup> To determine if replication fork protection may be mediated by ATR<sub>X</sub>-DAXX loading of histone H3.3, we performed a DNA fiber assay after depleting Daxx protein expression using a targeted small interfering RNA (siRNA) (siDAXX). DAXX depletion did not affect ATR<sub>X</sub> protein levels (Supplementary Figure 12C), but did have a significant effect on DNA tract length (Figures 5c and d). Pertaining to processivity, tracts from siDAXX-treated cells without HU were shorter than those from siScram control cells; however, as with the ATR<sub>X</sub> KD, HU-induced fork stalling resulted in significantly shorter labeled tracts. These findings

are consistent with a role for both ATRX and DAXX in the regulation of both replication fork processivity and protection upon fork stalling.

## Discussion

Neuronal progenitor cells of the ventricular (VZ)/subventricular (SVZ) zones sequentially exit the cell cycle to populate the

distinct neuronal layers of the forebrain. Inherently, the most proliferative neural progenitor cells (NPCs) that become the upper neuronal layers have the greatest potential to incur replication-induced DNA damage and subsequent genomic instability. In this regard, we demonstrated that *Atrx* deletion *in vivo* in NPCs specifically compromised the genesis of cells targeted for the upper neocortical layers (Figures 1b and 2e). At the molecular level, we demonstrate that ATRX is required



to diminish DNA replication stress, by protecting stalled replication forks, thereby preventing genomic damage and cell loss. Collectively, we propose a model in which ATRX is critical for heterochromatin maintenance throughout the cell cycle (Figure 6).

**ATRX and DAXX function to maintain heterochromatin stability.** Simple repeats are poor substrates for nucleosome recycling during DNA replication and represent regions of latent epigenomic instability.<sup>45–48</sup> Heterochromatin environments are essential for the preservation of structural elements, such as centromeres and telomeres, as well as for the repression of malicious DNA sequences encoding endogenous retroviral elements. The ATRX-DAXX histone chaperone deposits H3.3 at globally diffuse heterochromatic loci, including telomeres, centromeres, differentially methylated regions, CpG islands, and endogenous retroviral elements in a replication-independent manner.<sup>9,11,44,49</sup> Accordingly, the loss of ATRX leads to the dysregulation of these loci<sup>9,16,49</sup> and therefore we proposed a replication-independent mechanism for ATRX and DAXX to establish and maintain heterochromatin (Figure 6a). Although ATRX can recognize both HP1 and H3K9me3,<sup>4,50,51</sup> its H3.3 chaperone function appears to be upstream of SUV39H-mediated H3K9 trimethylation.<sup>9,49</sup> Additionally, ATRX's ability to bind to G4 structured DNA *in vitro*, as well as its high binding enrichment at G4 motif containing DNA sequences *in vivo*,<sup>17</sup> elicits the possibility that ATRX may recruit DAXX and H3.3 to G4 structured DNA for localized heterochromatinization (Figure 6a). Regardless, further experimentation is required to validate a role for ATRX in re-establishing heterochromatin, similar to studies identifying a role for Asf1 in histone recycling.<sup>52</sup> Importantly, G4 structured DNA can cause replication fork stalling, necessitating its suppression before the S phase,<sup>21</sup> whereas other studies have demonstrated fluid replication, although G4 motif DNA is required for the preservation of distinct epigenomic loci.<sup>46,47</sup>

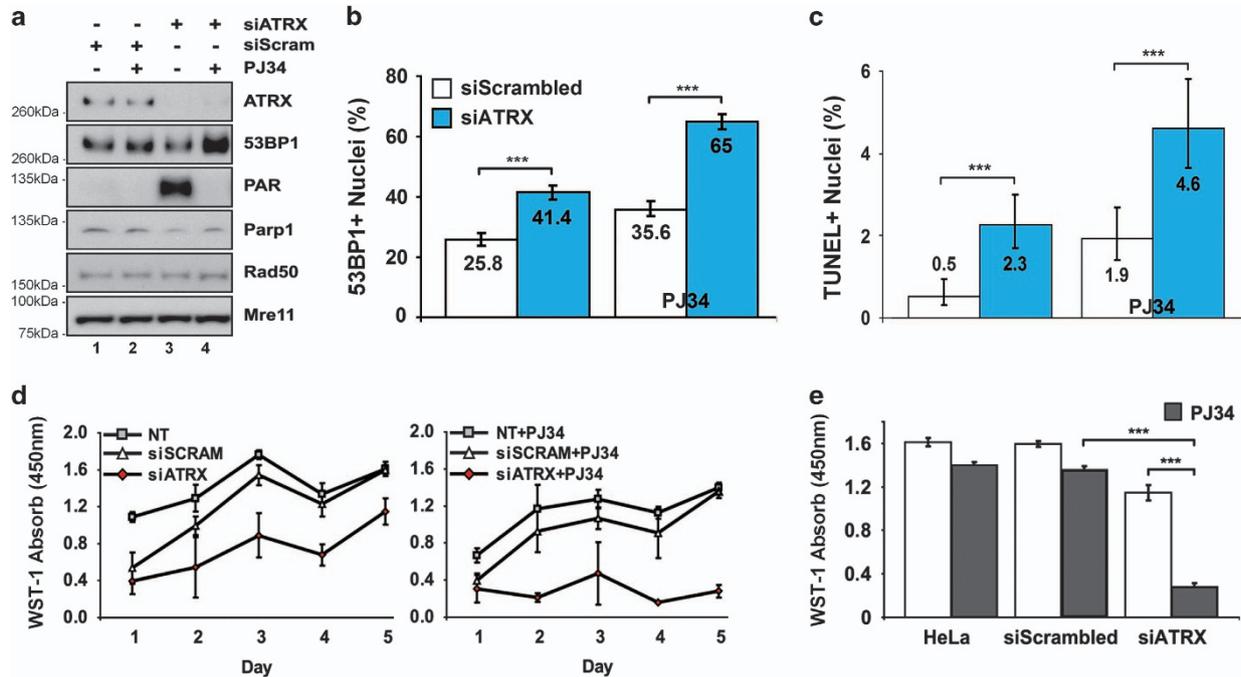
**ATRX actively protects stalled replication forks.** Here we progress our model into the S phase and propose a mechanism wherein ATRX actively protects stalled replication forks within heterochromatin (Figure 6b). ATRX-deficient cells are burdened by increased replication fork stalling events,<sup>15,20</sup> which are subsequently degraded by MRE11 (Figures 5a and b) in a manner akin to BRCA1/2-deficient cells (Supplementary Figure 11).<sup>30,53</sup> Adapting a previous model for ATRX regarding telomere maintenance,<sup>54</sup> we

propose that ATRX physically sequesters MRE11 to inhibit its exonuclease activity, thereby preventing fork degradation. BRCA1 co-localization with RAD51 marks the protection of stalled replication forks,<sup>29</sup> and we observed an increase in BRCA1 foci formation without a concomitant increase in RAD51 colocalization in ATRX-deficient cells. Unfettered MRE11 activity with an increased number of stalled replication forks may deplete RAD51 pools, and this may further attenuate stalled fork protection. In fact, a similar model has been proposed wherein ATR inhibition promoted precocious restart of stalled replication forks, thereby depleting RPA protein levels and ultimately leading to fork collapse.<sup>55</sup> Alternatively, dysregulated heterochromatin proximal to G4 structured DNA may cause ineffective mobilization of homologous recombination factors such as RAD51 in ATRX-deficient cells.

Furthermore, we propose that the upregulation of PARP-1 activity (Figures 2c and 4a) can be attributed to a compensatory mechanism that engages to protect stalled replication forks from MRE11-dependent degradation by PARP-1-mediated replication fork reversal<sup>35,56</sup> (Figure 6b). In this manner, the excessive processing of replicating heterochromatin in ATRX-null cells likely contributes to delayed S-phase progression (Supplementary Figure 5B; Clynes *et al.*<sup>15</sup> and Huh *et al.*<sup>16</sup>). Therefore, unresolved replication intermediates become DSBs in the subsequent G2 phase,<sup>57</sup> which may explain the increased DNA damage observed throughout the cell cycle (Figures 3b–d and 6c).

**Heterochromatin instability drives ATRX-associated disease.** Collectively, our data and others' suggests that enhanced cell death and reduced tissue size occurs from an inability to faithfully replicate heterochromatin under periods of rampant proliferation. The replication intermediates lead to DSBs, genomic instability, and mitotic catastrophe that reduces cell number (Figure 6c). Paradoxically, ATRX loss in cancer is beneficial to cell survival through the promotion of the alternative lengthening of telomere (ALT) phenotype. In this regard, ATRX loss is believed to be a late event, presumably after sufficient growth control checkpoints are eliminated. The instability of telomeric heterochromatin in the absence of ATRX facilitates telomere sister chromatid exchange, which maintains telomere length in ALT. Conversely, reintroduction of ATRX into ATRX-null ALT cancer cells restores H3.3 deposition at telomeres, thereby inhibiting sister telomere exchange and causing growth suppression.<sup>54</sup> Thus, our finding that small-molecule inhibition of PARP-1

**Figure 3** ATRX KD cells have increased activation of p53-ATM checkpoint upon mitotic progression and impaired RAD51 colocalization to BRCA1 foci. (a) Representative micrographs of phosphorylated ATM<sup>Ser1981</sup> (pATM; green) and cyclin A (CcnA; red) double IF staining of siScram- and siATRX-transfected HeLa cells at 96 h post-transfection. Arrowheads point to cells with DNA-damage foci. (b) Percentage of total interphase nuclei containing pATM foci in siATRX- versus siScram-transfected HeLa cells at 72 and 96 h post-transfection. siATRX: 72 h, *n* = 1001; 96 h, *n* = 1007. siScram: 72 h, *n* = 999; 96 h, *n* = 1009. (c) Percentage of S-G2 (CcnA+) nuclei containing pATM foci at 72 and 96 h post-transfection. siATRX: 72 h, *n* = 365; 96 h, *n* = 366. siScram: 72 h, *n* = 342; 96 h, *n* = 308. (d) Percentage of G1 (CcnA –) nuclei containing pATM foci at 72 and 96 h post-transfection. siATRX: 72 h, *n* = 636; 96 h, *n* = 641. siScram: 72 h, *n* = 657; 96 h, *n* = 701. (e) Representative micrographs of BRCA1 and RAD51 double immunostaining in siScram control and siATRX KD HeLa nuclei 72 h post-transfection. Solid arrowheads point to foci that are BRCA1+ and RAD51+ and open arrowheads point to foci that are only BRCA1+. (f) Scatterplot distribution profile of BRCA1 foci from the experiment described in (e). siScram, *n* = 106 nuclei; siATRX, *n* = 111 nuclei. (g) Quantification of BRCA1 foci from the experiment described in (e). siScram, *n* = 106 nuclei; siATRX, *n* = 111 nuclei. (h) Percentage of total BRCA1 foci positive for RAD51 from the experiment described in (e). All images are at ×630 magnification; scale bars are 20 μm (a) or 10 μm (e). For graphs, values represent percent total ± 95% CI except for (g), which is mean the number of BRCA1 foci ± S.E.M.; \*\**P* < 0.01; \*\*\**P* < 0.001 by z-scores (b–d and h) or Student's *t*-test (g)



**Figure 4** PARP-1 inhibition induces DNA breaks and causes growth suppression in ATRX KD cells. (a) Immunoblot analysis of PARP-1 inhibition by PJ34 in ATRX KD HeLa cells. As indicated, HeLa cells were transfected with siScram and siATRX. At 48 h after transfection, cells were treated with 5  $\mu$ M of PARP-1 inhibitor PJ34 (+) or untreated (-) for another 24 h. Whole-cell extracts were harvested 72 h post-transfection. (b) Percentage of total nuclei containing  $\geq 5$  bright 53BP1 foci in siScram- versus siATRX-transfected HeLa cells at 96 h post-transfection. At 72 h after transfection, cells were treated with 5  $\mu$ M of PARP-1 inhibitor PJ34 (right) or untreated for another 24 h (left). Cells were fixed 96 h post-transfection and stained for 53BP1. Values represent percent total  $\pm$  95% CI. siScram ( $n = 1420$ ); siATRX ( $n = 1607$ ); siScram+PJ34 ( $n = 1473$ ); siATRX+PJ34 ( $n = 1492$ ). \*\*\* $P < 0.001$  by z-scores. (c) Percentage of total nuclei containing TUNEL+ apoptotic nuclei in siScram- versus siATRX-transfected HeLa cells at 72 h post-transfection. At 48 h after transfection, cells were treated with 5  $\mu$ M of PARP-1 inhibitor PJ34 (right) or untreated for another 24 h (left). Cells were fixed 72 h post-transfection and TUNEL stained. Values represent percent total  $\pm$  95% CI. siScram ( $n = 2251$ ); siATRX ( $n = 2031$ ); siScram+PJ34 ( $n = 1802$ ); siATRX+PJ34 ( $n = 1455$ ). \*\*\* $P < 0.001$  by z-scores. (d) WST-1 cell viability time course of untreated (NT), siScram-, siATRX-transfected HeLa cells. Cells were seeded equally 24 h following transfection (left panel) or treated with 5  $\mu$ M PJ34 24 h later (right panel). Viability measurements were assessed at day 1 (72 h post-transfection) until day 5. Values represent mean  $\pm$  S.E.M. For all conditions,  $n = 4$ . (e) WST-1 cell viability measurement at day 5 of time courses described in (d)

activity attenuated growth of ATRX-deficient cells offers a potentially therapeutic avenue towards treatment of ALT-positive cancers, analogous to PARP-1 inhibitor treatment to eliminate BRCA1/2-deficient cancer cells.<sup>58–60</sup>

## Materials and Methods

**Animal husbandry.** *Atrx* cKO mice were generated by crossing ATRX floxed females (*ATRX*<sup>fl/fl</sup>) to *ATRX*<sup>+/Y</sup>:*FoxG1-Cre*<sup>+/+</sup> males on a C57BL/6 background as described previously.<sup>12</sup> *ATRX*<sup>fl/y</sup>:*FoxG1-Cre*<sup>+/+</sup> and *ATRX*<sup>fl/y</sup> (control) male littermates were harvested for analysis. Animal experiments were approved by the University of Ottawa's Animal Care ethics committee as per the guidelines set out by the Canadian Council on Animal Care.

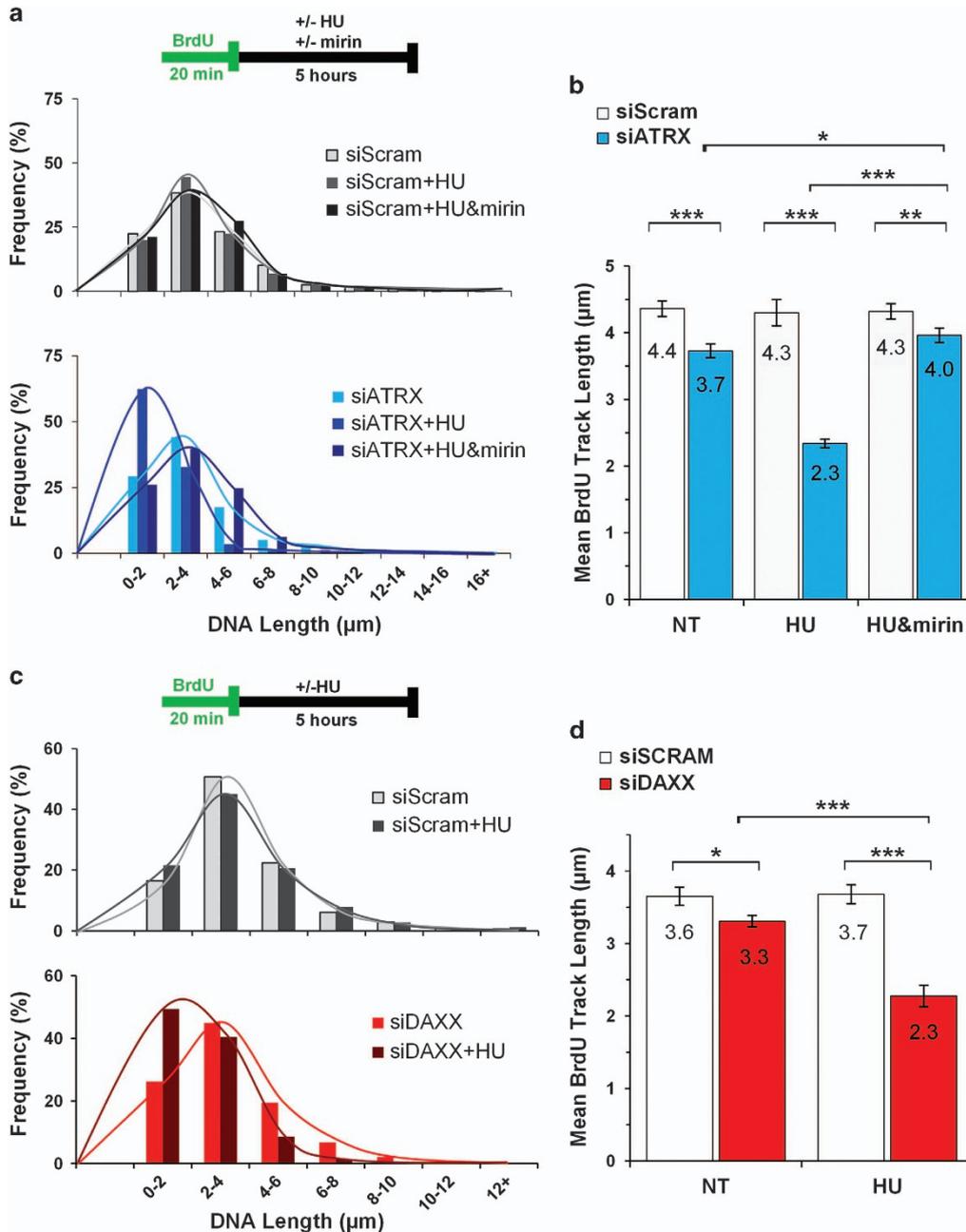
**Generation of ATRX shRNA cell lines.** The expression vector psiRNA-hH1neo (InvivoGen, Sand Diego, CA, USA) was digested with *Bbs1* and purified for cloning the ATRX shRNA oligonucleotide. The ATRX sense (5'-ACCTAACACTC ATCAGAAGAATCTGACCCTCAGATTCCTCTGATGAGTGTTC-3') and antisense (5'-CAAAAAACACTCATCAGAAGAATCTGAGGTGGTCAGATTCCTCTGATGAGTGTTC-3') oligonucleotides were designed with *Bbs1* overhangs. The oligonucleotides (25  $\mu$ M) were annealed in 150 mM NaCl by heating to 80  $^{\circ}$ C for 2 min followed by slow cooling to 37  $^{\circ}$ C. Annealed oligonucleotides were then ligated and cloned into the psiRNA-hH1neo plasmid. Recombinants were identified by an *Asel* digestion, purified using a Qiagen Maxiprep Kit (Qiagen, Toronto, ON, Canada), and sent for sequencing (StemCore, OHRI, Ottawa, ON, Canada). To generate stable cell lines, HeLa cells ( $5 \times 10^7$ ) were transfected with psiRNA expressing vectors by Lipofectamine (Life Technologies, Burlington, ON, Canada) as per the manufacturer's instructions. Clones were selected in DMEM supplemented with 800  $\mu$ g/ml

G418 (Life Technologies) after 2 weeks in culture. Individual clones were isolated and KD of ATRX protein expression was determined by western blot.

**Cell culture.** HeLa cells were cultured at 37  $^{\circ}$ C in DMEM with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin. Transient KD of ATRX and BRCA1 were performed on 50% confluent cells using 0.72% (v/v) INTERFERin (Polyplus, Illkirch, France) in Opti-MEM I Reduced Serum Medium (Thermo Fisher Scientific Inc., Waltham, MA, USA) as per the manufacturer's instructions, with 100 nM of either siATRX Smart Pool or a Scrambled control (GE Healthcare, Amersham, The Netherlands). siBRCA1 was a kind gift from Dr Christine Pratt (University of Ottawa, Ottawa, ON, Canada). PARP-1 was inhibited with 5  $\mu$ M PARP-1 inhibitor VIII (PJ34; Santa Cruz Biotechnology Inc.; sc-204161A).

For stable shRNA expressing clone growth curves, WT HeLa cells, psiRNA LacZ, and psiRNA ATRX-stable clones were G1 synchronized by 72 h serum withdrawal. Growth media were reintroduced at time 0 and cells were enumerated at the indicated time points.

**Protein extraction and immunoblot analysis.** Cortical lysates were extracted by homogenization using the Tissue Tearor (Biospec Products Inc., Bartlesville, OK, USA) in RIPA buffer (1 $\times$  PBS, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate, protease inhibitor Complete Mini EDTA-free in ddH<sub>2</sub>O). Cell culture lysates were extracted in RIPA buffer by gentle agitation. Protein samples were cleared by centrifugation at 4  $^{\circ}$ C and supernatants were quantified using the Bio-Rad Protein Assay reagent (Bio-Rad, Mississauga, ON, Canada). Protein samples were resolved on pre-cast 3–8% Tris-acetate or 4–12% Tris-Bis gels (NuPage; Life Technologies) and transferred onto PVDF membrane (Immobilon-P; Millipore, Etobicoke, ON, Canada). Membranes were probed with the indicated primary antibodies (see Supplementary Table S1) and HRP-conjugated secondary

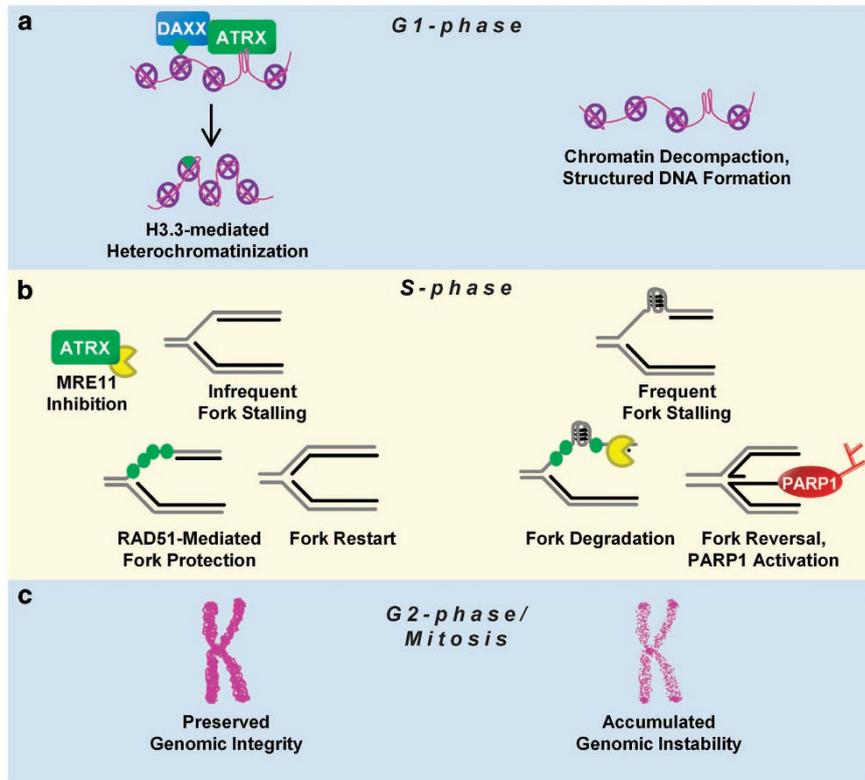


**Figure 5** The ATRX-DAXX pathway protects stalled DNA replication forks from degradation by MRE11 exonuclease activity. (a) DNA fiber tract length distribution histogram of siScram- (top) and siATRAX- (bottom) transfected HeLa cells at 72 h post-transfection. siRNA-treated cells were pulsed with BrdU and subsequently exposed to HU and mirin as indicated in the schematic. Total fibers counted for siScram experiment: no treatment, NT ( $n=1782$ ); HU ( $n=1819$ ); HU and mirin ( $n=1759$ ). Total fibers counted for siATRAX-treated cells: NT ( $n=1527$ ); HU ( $n=1523$ ); HU and mirin ( $n=1536$ ). (b) Mean DNA fiber tract length of experiments described in (a). (c) DNA fiber tract length distribution histogram of siScram- (top) and siDAXX- (bottom) transfected HeLa cells at 72 h post-transfection. Fibers counted for siScram-treated cells were: NT ( $n=888$ ); HU ( $n=998$ ). Total fibers counted for siDAXX-treated cells were: NT ( $n=888$ ) and HU ( $n=1171$ ). (d) Mean DNA fiber tract length of experiments described in (c). For panels (b and d), the mean length  $\pm$  95% CI was plotted. \* $P < 0.05$ , \*\* $P < 0.01$  and \*\*\* $P < 0.001$  by Mann-Whitney test

antibodies. Immunoblots were incubated with enhanced chemiluminescent substrate and signals were exposed to film. Densitometric gel analysis was performed using ImageJ (version 1.46j; Bethesda, MA, USA) software by integrating pixel density plots with background subtraction.

**IF microscopy for cell culture.** Cells were grown on coverslips or cytospun (Cytospin 4 Cytocentrifuge; Thermo Fisher Scientific Inc.) onto slides and fixed in

2% PFA and permeabilized with 0.1% Triton-X. Primary antibodies (see Supplementary Table S1) were diluted in blocking buffer (20% horse serum, 0.1% FBS, 0.03% sodium azide, in PBS) and incubated overnight at 4 °C in a humidifying chamber. Secondary antibodies (Alexas 488 and 594; Life Technologies) were applied and nuclei were counterstained with DAPI. Images were taken with an Axio Imager M1 microscope (Zeiss, Toronto, ON, Canada) and analyzed using ImageJ software. Positively stained cells were scored as indicated, relative to DAPI-stained nuclei.



**Figure 6** A model of how ATRX suppresses genomic instability during cellular proliferation. Relevant scenarios are shown during (a) G1 phase, (b) S phase and (c) G2/M phase in the presence (left) or absence (right) of ATRX. (a) During G1, ATRX localizes to decompacted and structured DNA (e.g. G4-DNA) along with DAXX to chaperone H3.3-H4 dimers that serve as a beacon for further heterochromatinization. When cells progress into the S phase (b), DNA replication forks experience more frequent stalling events when ATRX is absent owing to an increased incidence of structured DNA. ATRX physically interacts with MRE11 and inhibits excessive MRE11-mediated resectioning of stalled replication forks, which subsequently require RAD51-mediated protection of nascent DNA. In the absence of ATRX, PARP-1 activation is upregulated in an attempt to reverse stalled replication forks and protect against further MRE11 resectioning. Cells with frequent fork stalling that progress into the G2 phase and mitosis (c) are more prone to DSBs and mutagenic non-allelic homologous recombination events (NAHR) resulting in genomic instability

**IF microscopy for brain sections.** Embryos were harvested at the indicated gestational time points. Heads from embryos were fixed in 4% PFA overnight at 4 °C. The heads were washed in PBS, cryoprotected in a 30% sucrose/PBS solution overnight at 4 °C, embedded in a 1:1 solution of 30% sucrose and OCT Compound (Tissue-Tek), and flash frozen on liquid nitrogen. Embedded tissue were serially sectioned at 10  $\mu$ m (Leica 1850 cryostat) and mounted onto Superfrost Plus-coated slides (Thermo Fisher Scientific) and dried at room temperature for 2 h. Slides were fixed with 70% ethanol for 5 min at 4 °C (IHC) or 2% PFA 10 min at room temperature and then rehydrated in 1 $\times$  PBS for 5 min before staining. When probing for PAR, slides were incubated in 2 N HCl for 20 min at 37 °C. Sections were permeabilized (0.1% Tween-20, 0.1M Tris-HCl (pH 8.8)) and incubated in blocking buffer (20% goat serum, 0.3% Triton-X in PBS). Primary antibodies (see Supplementary Table S1) were diluted in blocking buffer and applied onto sections. Sections were washed in PBS, incubated in secondary antibody solution, and counterstained with DAPI. Images were taken with an Axio Imager M1 microscope (Zeiss). Marker-positive cell counts were performed on multiple ( $n > 3$ ) 200  $\mu$ m brain sections from the dorsal cortex and plotted as a percentage of the total number of DAPI-positive cells.

**Cell cycle progression analysis.** HeLa psiRNA LacZ and psiRNA ATRX cells were pulsed with 30  $\mu$ M BrdU containing media in triplicate for each time point (0, 6, 12, 16, 20, 24, and 28 h). A total of 10<sup>6</sup> cells were fixed with 1 ml of 70% ethanol solution at -20 °C, overnight, resuspended in 0.1 N HCl+0.7% Triton-X on ice for 15 min, and washed with PBS. Cells were stained in 1:100 dilution the primary antibody anti-BrdU (BD Biosciences, San Jose, CA, USA) diluted in HBT (PBS, 0.05% FBS, 0.005% Tween-20), washed with HBT, and stained with FITC-conjugated secondary antibody anti-mouse diluted 1:20 in HBT for 30 min in the dark and precipitated for 7 min at 1500 r.p.m. Cells were resuspended in PI

(propidium iodide) solution with RNase A (50  $\mu$ g/ml PI, 40  $\mu$ g/ml RNase A) at 2000 cells per  $\mu$ l and analyzed by flow cytometry using a Beckman Coulter FACS station (Brea, CA, USA). Cell cycle distribution of the cell population was analyzed with the FCS Express 2 software (DeNovo Software, Thornhill, ON, Canada) and the cell-cycle profile of each time point was analyzed with the ModFit software (Verity Software House, Topsham, ME, USA).

**Caspase assays.** Cortical lysate protein was added to freshly prepared caspase activity buffer (25 mM HEPES, 10% sucrose, 1 mM EDTA, 0.1% CHAPS, 10 mM DTT in ddH<sub>2</sub>O) for a total volume of 199  $\mu$ l per well. The reaction was initiated by the addition of 1  $\mu$ l of 10 mM fluorescent substrate (caspase-3 substrate, Ac-DEVD-AMC (P411; Biomol, Hamburg, Germany), caspase-8 substrate, Ac-IETD-AMC (P432; Biomol, Hamburg, Germany), caspase-9 substrate, Ac-LEHD-AMC (P444; Biomol, Hamburg, Germany)) to each well. A Thermo-Labsystems Fluoroskan Ascent FL fluorometer using an excitation filter set to 380 nm and an emission filter set to 460 nm was used to read the absorbance of each well every 5 min over a 2 h period.

**DNA fiber assay.** Nascent DNA of HeLa cells treated with siATRX, siBRCA1 or siScram was labeled with a 50  $\mu$ M BrdU pulse and replication forks were stalled with 4 mM HU. Where indicated, cells were treated with the MRE11 inhibitor mirin at a concentration of 50  $\mu$ M. A total of 10<sup>6</sup> cells per 2  $\mu$ l were spotted onto glass slides and lysed with 7  $\mu$ l of fiber lysis solution (50 mM EDTA, 0.5% SDS and 200 mM Tris-HCl) for 5 min at RT. Slides were tilted 15° to horizontal to spread DNA across the length of the slide, and then air-dried and fixed in methanol/acetic acid (3:1). Slides were immersed in 2.5 N HCl for 80 min, washed in PBS, blocked in 5% BSA and stained with 1:500 mouse anti-BrdU primary antibody (BD Biosciences), followed by 1:4000 donkey anti-mouse IgG (H+L) with Alexa Fluor 488 conjugate

secondary antibody (Thermo Fisher Scientific Inc.). Indicated numbers of labeled DNA fibers from three independent experiments per condition were imaged (Zeiss Axio Imager M1 microscope, Oberkochen, Germany) and analyzed using the ImageJ software (Bethesda, MD, USA).

**WST assay.** WST-1 proliferation assay was performed as per the manufacturer's instructions (ab65473; Abcam, Cambridge, UK). HeLa cells were seeded at 1000 cells per well on a 96-well plate and absorbance measured at 450nm.

**TUNEL assay.** Cells were fixed with 2% PFA and were permeabilized in 0.1% Triton-X/0.1% sodium citrate for 2 min on ice. The TUNEL labeling was performed using the *In Situ* Cell Death Detection Kit (Roche Applied Science, Mississauga, ON, Canada) according to the manufacturer's instructions.

**Statistical analysis.** Statistical analysis was performed using Microsoft Excel statistical analysis package with means and s.e. calculated. Significance was determined by two-tailed *t*-tests of unequal variance (95 and 99% confidence intervals). Additionally, *P*-values for fiber assays were determined by Mann–Whitney test. All significant *P*-values were marked with asterisks, as follows: \**P*<0.05, \*\**P*<0.01, and \*\*\**P*<0.001.

### Conflict of Interest

The authors declare no conflict of interest.

**Acknowledgements.** We thank Jeff Hamill and Dr. Bruce McKay for assistance with FACS analysis. This work was supported by operating grants to DJP from the Canadian Institutes of Health Research (MOP-133586) and from the Cancer Research Society, University of Ottawa, and Ottawa Hospital Research Institute partnership program. DI was supported by a CIHR CGS-M award.

1. Kleefstra T, Schenck A, Kramer JM, van Bokhoven H. The genetics of cognitive epigenetics. *Neuropharmacology* 2014; **80**: 83–94.
2. Gibbons RJ, Higgs DR. Molecular-clinical spectrum of the ATR-X syndrome. *Am J Med Genet* 2000; **97**: 204–212.
3. Gibbons RJ, Picketts DJ, Villard L, Higgs DR. Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). *Cell* 1995; **80**: 837–845.
4. Iwase S, Xiang B, Ghosh S, Ren T, Lewis PW, Cochrane JC et al. ATRX ADD domain links an atypical histone methylation recognition mechanism to human mental-retardation syndrome. *Nat Struct Mol Biol* 2011; **18**: 769–776.
5. Noh KM, Maze I, Zhao D, Xiang B, Wenderski W, Lewis PW et al. ATRX tolerates activity-dependent histone H3 methyl/phos switching to maintain repetitive element silencing in neurons. *Proc Natl Acad Sci USA* 2014; **112**: 6820–6827.
6. Picketts DJ, Higgs DR, Bachoo S, Blake DJ, Quarrell OW, Gibbons RJ. ATRX encodes a novel member of the SNF2 family of proteins: mutations point to a common mechanism underlying the ATR-X syndrome. *Hum Mol Genet* 1996; **5**: 1899–1907.
7. Lewis PW, Elsaesser SJ, Noh KM, Stadler SC, Allis CD. Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proc Natl Acad Sci USA* 2010; **107**: 14075–14080.
8. Tang J, Wu S, Liu H, Stratt R, Barak OG, Shiekhattar R et al. A novel transcription regulatory complex containing Daxx and the ATR-X syndrome protein. *J Biol Chem* 2004; **279**: 20369–20377.
9. Voon HP, Hughes JR, Rode C, De La Rosa-Velázquez IA, Jenuwein T, Feil R et al. ATRX plays a key role in maintaining silencing at interstitial heterochromatic loci and imprinted genes. *Cell Rep* 2015; **11**: 405–418.
10. Wong LH, McGhie JD, Sim M, Anderson MA, Ahn S, Hannan RD et al. ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. *Genome Res* 2010; **20**: 351–360.
11. Elsasser SJ, Noh KM, Diaz N, Allis CD, Banaszynski LA. Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature* 2015; **522**: 240–244.
12. Berube NG, Mangelsdorf M, Jagla M, Vanderluit J, Garrick D, Gibbons RJ et al. The chromatin-remodeling protein ATRX is critical for neuronal survival during corticogenesis. *J Clin Invest* 2005; **115**: 258–267.
13. Gibbons R. Alpha thalassaemia-mental retardation, X linked. *Orphanet J Rare Dis* 2006; **1**: 15.
14. Bagheri-Fam S, Argentaro A, Svingen T, Combes AN, Sinclair AH, Koopman P et al. Defective survival of proliferating Sertoli cells and androgen receptor function in a mouse model of the ATR-X syndrome. *Hum Mol Genet* 2011; **20**: 2213–2224.
15. Clynes D, Jelińska K, Xella B, Ayyub H, Taylor S, Mitson M et al. ATRX dysfunction induces replication defects in primary mouse cells. *PLoS One* 2014; **9**: e92915.

16. Huh MS, Price O'Dea T, Ouazia D, McKay BC, Parise G, Parks RJ et al. Compromised genomic integrity impedes muscle growth after Atrx inactivation. *J Clin Invest* 2012; **122**: 4412–4423.
17. Law MJ, Lower KM, Voon HP, Hughes JR, Garrick D, Viprakasit V et al. ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell* 2010; **143**: 367–378.
18. Clynes D, Gibbons RJ. ATRX and the replication of structured DNA. *Curr Opin Genet Dev* 2013; **23**: 289–294.
19. Clynes D, Higgs DR, Gibbons RJ. The chromatin remodeller ATRX: a repeat offender in human disease. *Trends Biochem Sci* 2013; **38**: 461–466.
20. Leung JW, Ghosal G, Wang W, Shen X, Wang J, Li L et al. Alpha thalassemia/mental retardation syndrome X-linked gene product ATRX is required for proper replication restart and cellular resistance to replication stress. *J Biol Chem* 2013; **288**: 6342–6350.
21. Mirkin EV, Mirkin SM. Replication fork stalling at natural impediments. *Microbiol Mol Biol Rev* 2007; **71**: 13–35.
22. Petermann E, Orta ML, Issaeva N, Schultz N, Helleday T. Hydroxyurea-stalled replication forks become progressively inactivated and require two different RAD51-mediated pathways for restart and repair. *Mol Cell* 2010; **37**: 492–502.
23. Palazzo L, Della Monica R, Visconti R, Costanzo V, Grieco D. ATM controls proper mitotic spindle structure. *Cell Cycle* 2014; **13**: 1091–1100.
24. Yang YG, Cortes U, Patnaik S, Jasin M, Wang ZQ. Ablation of PARP-1 does not interfere with the repair of DNA double-strand breaks, but compromises the reactivation of stalled replication forks. *Oncogene* 2004; **23**: 3872–3882.
25. Hunt T, Luca FC, Ruderman JV. The requirements for protein synthesis and degradation, and the control of destruction of cyclins A and B in the meiotic and mitotic cell cycles of the clam embryo. *J Cell Biol* 1992; **116**: 707–724.
26. Pines J, Hunter T. Human cyclins A and B1 are differentially located in the cell and undergo cell cycle-dependent nuclear transport. *J Cell Biol* 1991; **115**: 1–17.
27. Peng JC, Karpen GH. Epigenetic regulation of heterochromatic DNA stability. *Curr Opin Genet Dev* 2008; **18**: 204–211.
28. McDowell TL, Gibbons RJ, Sutherland H, O'Rourke DM, Bickmore WA, Pombo A et al. Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes. *Proc Natl Acad Sci USA* 1999; **96**: 13983–13988.
29. Feng Z, Zhang J. A dual role of BRCA1 in two distinct homologous recombination mediated repair in response to replication arrest. *Nucleic Acids Res* 2012; **40**: 726–738.
30. Schlacher K, Wu H, Jasin M. A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2. *Cancer Cell* 2012; **22**: 106–116.
31. Scully R, Chen J, Ochs RL, Keegan K, Hoekstra M, Feunteun J et al. Dynamic changes of BRCA1 subnuclear location and phosphorylation state are initiated by DNA damage. *Cell* 1997; **90**: 425–435.
32. Scully R, Chen J, Plug A, Xiao Y, Weaver D, Feunteun J et al. Association of BRCA1 with Rad51 in mitotic and meiotic cells. *Cell* 1997; **88**: 265–275.
33. Bryant HE, Petermann E, Schultz N, Jemth AS, Loseva O, Issaeva N et al. PARP is activated at stalled forks to mediate Mre11-dependent replication restart and recombination. *EMBO J* 2009; **28**: 2601–2615.
34. Hassler M, Ladurner AG. Towards a structural understanding of PARP1 activation and related signalling ADP-ribosyl-transferases. *Curr Opin Struct Biol* 2012; **22**: 721–729.
35. Ying S, Hamdy FC, Helleday T. Mre11-dependent degradation of stalled DNA replication forks is prevented by BRCA2 and PARP1. *Cancer Res* 2012; **72**: 2814–2821.
36. De Lorenzo SB, Patel AG, Hurley RM, Kaufmann SH. The elephant and the blind men: making sense of PARP inhibitors in homologous recombination deficient tumor cells. *Front Oncol* 2013; **3**: 228.
37. Gottipati P, Viscioni B, Schultz N, Solomons J, Bryant HE, Djureinovic T et al. Poly(ADP-ribose) polymerase is hyperactivated in homologous recombination-defective cells. *Cancer Res* 2010; **70**: 5389–5398.
38. Ame JC, Spenlehauer C, de Murcia G. The PARP superfamily. *BioEssays* 2004; **26**: 882–893.
39. Steffen JD, Brody JR, Armen RS, Pascal JM. Structural implications for selective targeting of PARPs. *Front Oncol* 2013; **3**: 301.
40. Petermann E, Helleday T. Pathways of mammalian replication fork restart. *Nat Rev Mol Cell Biol* 2010; **11**: 683–687.
41. Hashimoto Y, Ray Chaudhuri A, Lopes M, Costanzo V. Rad51 protects nascent DNA from Mre11-dependent degradation and promotes continuous DNA synthesis. *Nat Struct Mol Biol* 2010; **17**: 1305–1311.
42. Dupre A, Boyer-Chatenet L, Sattler RM, Modi AP, Lee JH, Nicolette ML et al. A forward chemical genetic screen reveals an inhibitor of the Mre11-Rad50-Nbs1 complex. *Nat Chem Biol* 2008; **4**: 119–125.
43. Frey A, Listovsky T, Guilbaud G, Sarkies P, Sale JE. Histone H3.3 is required to maintain replication fork progression after UV damage. *Curr Biol* 2014; **24**: 2195–2201.
44. Goldberg AD, Banaszynski LA, Noh KM, Lewis PW, Elsaesser SJ, Stadler S et al. Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* 2010; **140**: 678–691.
45. Sarkies P, Murat P, Phillips LG, Patel KJ, Balasubramanian S, Sale JE et al. FANCD1 coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. *Nucleic Acids Res* 2012; **40**: 1485–1498.

46. Sarkies P, Reams C, Simpson LJ, Sale JE. Epigenetic instability due to defective replication of structured DNA. *Mol Cell* 2010; **40**: 703–713.
47. Schiavone D, Guilbaud G, Murat P, Papadopoulou C, Sarkies P, Prioleau MN *et al*. Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. *EMBO J* 2014; **33**: 2507–2520.
48. Schneiderman JI, Sakai A, Goldstein S, Ahmad K. The XNP remodeler targets dynamic chromatin in *Drosophila*. *Proc Natl Acad Sci USA* 2009; **106**: 14472–14477.
49. He Q, Kim H, Huang R, Lu W, Tang M, Shi F *et al*. The Daxx/Atrx complex protects tandem repetitive elements during DNA hypomethylation by promoting H3K9 trimethylation. *Cell Stem Cell* 2015; **17**: 273–286.
50. Berube NG, Smeenk CA, Picketts DJ. Cell cycle-dependent phosphorylation of the ATRX protein correlates with changes in nuclear matrix and chromatin association. *Hum Mol Genet* 2000; **9**: 539–547.
51. Eustermann S, Yang JC, Law MJ, Amos R, Chapman LM, Jelinska C *et al*. Combinatorial readout of histone H3 modifications specifies localization of ATRX to heterochromatin. *Nat Struct Mol Biol* 2011; **18**: 777–782.
52. Jasencakova Z, Scharf AN, Ask K, Corpet A, Imhof A, Almouzni G *et al*. Replication stress interferes with histone recycling and predeposition marking of new histones. *Mol Cell* 2010; **37**: 736–743.
53. Schlacher K, Scharf AN, Ask K, Corpet A, Imhof A, Almouzni G *et al*. Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. *Cell* 2011; **145**: 529–542.
54. Clynes D, Jelinska C, Xella B, Ayyub H, Scott C, Mitson M *et al*. Suppression of the alternative lengthening of telomere pathway by the chromatin remodelling factor ATRX. *Nat Commun* 2015; **6**: 7538.
55. Toledo LI, Altmeyer M, Rask MB, Lukas C, Larsen DH, Povlsen LK *et al*. ATR prohibits replication catastrophe by preventing global exhaustion of RPA. *Cell* 2013; **155**: 1088–1103.
56. Ray Chaudhuri A, Hashimoto Y, Herrador R, Neelsen KJ, Fachinetti D, Bermejo R *et al*. Topoisomerase I poisoning results in PARP-mediated replication fork reversal. *Nat Struct Mol Biol* 2012; **19**: 417–423.
57. Bruhn C, Zhou ZW, Ai H, Wang ZQ. The essential function of the MRN complex in the resolution of endogenous replication intermediates. *Cell Rep* 2014; **6**: 182–195.
58. Lee JM, Ledermann JA, Kohn EC. PARP inhibitors for BRCA1/2 mutation-associated and BRCA-like malignancies. *Ann Oncol* 2014; **25**: 32–40.
59. Lord CJ, Tutt AN, Ashworth A. Synthetic lethality and cancer therapy: lessons learned from the development of PARP inhibitors. *Annu Rev Med* 2015; **66**: 455–470.
60. Xiong T, Wei H, Chen X, Xiao H. PJ34, a poly(ADP-ribose) polymerase (PARP) inhibitor, reverses melphalan-resistance and inhibits repair of DNA double-strand breaks by targeting the FA/BRCA pathway in multidrug resistant multiple myeloma cell line RPMI8226/R. *Int J Oncol* 2015; **46**: 223–232.



**Cell Death and Disease** is an open-access journal published by Nature Publishing Group. This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

Supplementary Information accompanies this paper on Cell Death and Disease website (<http://www.nature.com/cddis>)

# ATR-X Syndrome Protein Targets Tandem Repeats and Influences Allele-Specific Expression in a Size-Dependent Manner

Martin J. Law,<sup>1,8</sup> Karen M. Lower,<sup>1,8</sup> Hsiao P.J. Voon,<sup>1</sup> Jim R. Hughes,<sup>1</sup> David Garrick,<sup>1</sup> Vip Viprakasit,<sup>3</sup> Matthew Mitson,<sup>1</sup> Marco De Gobbi,<sup>1</sup> Marco Marra,<sup>7</sup> Andrew Morris,<sup>4</sup> Aaron Abbott,<sup>4</sup> Steven P. Wilder,<sup>5</sup> Stephen Taylor,<sup>2</sup> Guilherme M. Santos,<sup>6</sup> Joe Cross,<sup>1</sup> Helena Ayyub,<sup>1</sup> Steven Jones,<sup>7</sup> Jiannis Ragoussis,<sup>4</sup> Daniela Rhodes,<sup>6</sup> Ian Dunham,<sup>5</sup> Douglas R. Higgs,<sup>1</sup> and Richard J. Gibbons<sup>1,\*</sup>

<sup>1</sup>Medical Research Council Molecular Haematology Unit

<sup>2</sup>Computational Biology Research Group

Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DS, UK

<sup>3</sup>Department of Paediatrics, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand

<sup>4</sup>The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK

<sup>5</sup>European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

<sup>6</sup>Structural Studies Division, Medical Research Council, Laboratory of Molecular Biology, Cambridge CB2 0QH, UK

<sup>7</sup>BCCA Genome Sciences Centre, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

<sup>8</sup>These authors contributed equally to this work

\*Correspondence: richard.gibbons@imm.ox.ac.uk

DOI 10.1016/j.cell.2010.09.023

## SUMMARY

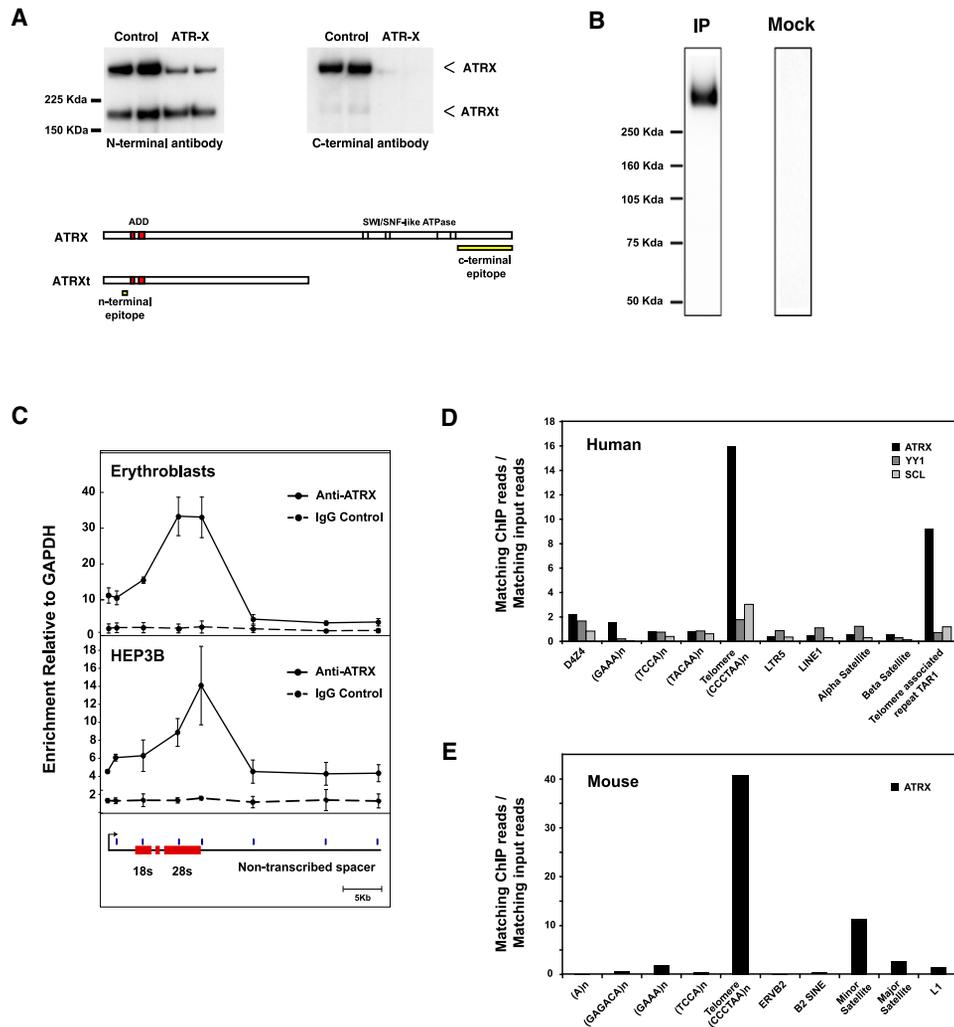
**ATR-X is an X-linked gene of the SWI/SNF family, mutations in which cause syndromal mental retardation and downregulation of  $\alpha$ -globin expression. Here we show that ATRX binds to tandem repeat (TR) sequences in both telomeres and euchromatin. Genes associated with these TRs can be dysregulated when ATRX is mutated, and the change in expression is determined by the size of the TR, producing skewed allelic expression. This reveals the characteristics of the affected genes, explains the variable phenotypes seen with identical ATRX mutations, and illustrates a new mechanism underlying variable penetrance. Many of the TRs are G rich and predicted to form non-B DNA structures (including G-quadruplex) in vivo. We show that ATRX binds G-quadruplex structures in vitro, suggesting a mechanism by which ATRX may play a role in various nuclear processes and how this is perturbed when ATRX is mutated.**

## INTRODUCTION

Although it is known that proteins of the Swi/Snf family are required to facilitate a wide range of nuclear processes (e.g., replication, recombination, repair, transcription), the mechanisms by which they operate in vivo are poorly understood (Flaus et al., 2006). One such widely expressed protein (ATRX) was first identified when it was shown that mutations in the X-linked gene (ATRX) caused a form of syndromal mental retardation, with

multiple developmental abnormalities characteristically associated with  $\alpha$  thalassaemia (ATR-X syndrome) (Gibbons et al., 1995). To date 127 disease-causing mutations have been found, most of which are located in two highly conserved domains of the ATRX protein (Gibbons et al., 2008). At the N terminus these lie within a globular domain (similar to that found in DNMT3 and DNMT3L, the so-called ADD domain) including a plant homeodomain (PHD), which most probably binds the N-terminal tails of histone H3 (Argentaro et al., 2007). At the C terminus there are seven helicase subdomains that identify ATRX as a member of the SNF2 family of chromatin-associated proteins (Figure 1A). Although many of these proteins have been shown to remodel, remove, or slide nucleosomes using in vitro assays, ATRX is most closely related to a subgroup (including RAD54 and ARIP4) that, despite acting as ATP-driven molecular motors, perform poorly in such canonical assays, suggesting that they have related but different chromatin-associated functions (Xue et al., 2003 and unpublished data).

Some clues to the role of ATRX in vivo have come from studying its distribution in the nucleus, the proteins with which it interacts, and the effects of mutations. Using indirect immunofluorescence, ATRX is found at heterochromatic repeats, at rDNA repeats, at telomeric repeats, and within PML bodies, which themselves are often associated with heterochromatic structures including telomeres (Gibbons et al., 2000; McDowell et al., 1999; Xue et al., 2003). Two robust protein-protein interactions have been described. The first occurs with DAXX (Xue et al., 2003) (a protein that is also found in PML bodies), which has been implicated in both pro- and antiapoptotic pathways. The second interaction occurs with HP1 $\alpha$  and HP1 $\beta$ , proteins that are widely associated with heterochromatin, including the telomere (Berube et al., 2000). It has also been shown that mutations in ATRX are consistently associated with alterations in the



**Figure 1. Validation of ATRX ChIP Protocol**

(A) Immunoblots of protein extracts from ATR-X patient and normal control lymphoblastoid cell lines using ATRX N- and C-terminal antibodies. The ATR-X patient harbors an ATRX C-terminal deletion mutation affecting the C-terminal antibody epitope. Schematic diagram of ATRX shows protein isoforms, antibody epitope regions, and conserved domains.

(B) The ATRX C-terminal antibody crosslinked to protein A-Sepharose was used to immunopurify ATRX from EBV cells. Eluted protein was analyzed by western blot probed with the N-terminal mouse monoclonal ATRX antibody, 39f. The mock control lane contains sample immunopurified using normal rabbit IgG.

(C) Q-PCR analysis of ATRX ChIP at the major ribosomal RNA gene locus in erythroblast (n = 4) and Hep3B (n = 3). Error bars show standard deviations. Diagram of the ribosomal RNA gene locus shows positions of rRNAs (red boxes), the promoter (arrow), and the Q-PCR primers (boxes above line).

(D) Direct mapping of human ATRX, SCL, and YY1 ChIP-seq reads to simple and interspersed repeats. Selected representative data are shown. For the complete dataset, see Table S1.

(E) Direct mapping of ATRX ChIP-seq sequence reads to mouse simple and interspersed repeats.

See Figure S1 for further validation of the specificity of the ATRX ChIP.

pattern of DNA methylation at such repeat sequences (rDNA, interstitial heterochromatic repeats, and subtelomeric repeats) (Gibbons et al., 2000).

Recently, an important link has been established between these observations and more functional studies. First, it has been shown that ATRX and HP1 localize to the telomeres of chromosomes in mouse embryonic stem cells (ESCs) (Wong et al., 2010). Second, it has been shown that ATRX localizes to telomeres in synchrony with the histone variant H3.3. Using immunoprecipitation it was shown that ATRX and its partner

DAXX specifically interact with H3.3, which is found to be associated with both active and inactive genes, regulatory elements, and telomeres (Goldberg et al., 2010). It has recently been shown that DAXX is an H3.3-specific chaperone (Drané et al., 2010; Lewis et al., 2010), and in the absence of ATRX, H3.3 is no longer recruited to telomeres whereas recruitment to the interstitial sites that were analyzed appeared to be unaffected (Goldberg et al., 2010). These observations suggest that ATRX plays an important role in establishing or maintaining the chromatin environment of telomeres and subtelomeric regions where it facilitates histone

replacement with the H3.3 variant (Drané et al., 2010; Lewis et al., 2010).

Although these observations have provided new insight into the potential role of ATRX at heterochromatic regions of the genome, they have not identified the euchromatic targets of ATRX and have not addressed the role of ATRX in regulating gene expression. To date the only human genes whose expression is known to be affected by ATRX mutations lie in the  $\alpha$ -globin gene cluster (Gibbons et al., 1991). Although clearly related to the  $\beta$ -globin cluster throughout evolution, ATRX mutations do not affect  $\beta$ -globin expression. It has been noted that the structure (e.g., GC content, repeat density, gene density), nuclear organization (e.g., nuclear position, relationship to chromosome territory, relationship to heterochromatin), and epigenetic environment (e.g., timing of replication, chromatin modification, DNA methylation) associated with these two clusters are radically different (Higgs et al., 1998). Most notably the human  $\alpha$ -globin cluster lies very close to the telomere of chromosome 16. It has previously been suggested that ATRX is targeted to specific regions of the genome defined by their genomic organization and/or chromatin structure. Thus mutations in ATRX may affect one type of chromosomal region (e.g., containing the  $\alpha$ -globin genes) but not another (e.g., containing the  $\beta$ -globin genes).

Here we have established the genome-wide distribution of the ATRX protein in both mouse and human cells. We have confirmed that ATRX binds directly to mouse telomeres and also shown that ATRX is enriched at the telomeres and subtelomeric regions of human chromosomes. Chromatin immunoprecipitation (ChIP) sequencing identified 917 targets in primary human erythroid cells (in which the globin genes are expressed) and 1305 targets in mouse ESCs. The most prominent feature of the targets in both human and mouse is the presence of variable number tandem repeats (VNTRs), which in many (but not all) cases are G and C rich and contain a high proportion of CpG dinucleotides. Of particular interest we show that, when ATRX function is compromised in ATR-X syndrome, the degree of perturbation in gene expression is related to the size of the TR, and this may lead to monoallelic expression. These findings explain the variable phenotypes seen in patients with identical ATRX mutations and provide a new mechanism underlying variable penetrance. A common theme shared by telomeres and many of the subtelomeric targets of ATRX is their potential to form G-quadruplex (G4) DNA structures. Here we show that ATRX binds G4 DNA *in vitro*, suggesting a common mechanism by which ATRX may influence a wide range of nuclear processes in the telomeric, subtelomeric, and interstitial regions of mammalian chromosomes.

## RESULTS

### Validation of an ATRX ChIP Protocol with rDNA as a Target

Domain structure, interaction partners, and biochemical activity currently implicate ATRX in the regulation of transcription via a physical interaction with chromatin. To date, ATRX has been implicated in histone H3.3 deposition at telomeres, but little is known about ATRX function away from telomeres because no direct ATRX target genes have been described. To address

this, an ATRX ChIP assay was developed using the ribosomal gene loci (rDNA) as the first candidate targets. The rDNA loci were chosen because immunofluorescence studies have previously shown that, in mitotic cells, ATRX is consistently found on the short arms of the acrocentric chromosomes in human colocalizing with the rDNA loci (McDowell et al., 1999); rDNA also becomes hypomethylated at CpG dinucleotides in primary peripheral blood mononuclear cells (PBMCs) from patients with ATR-X syndrome (Gibbons et al., 2000).

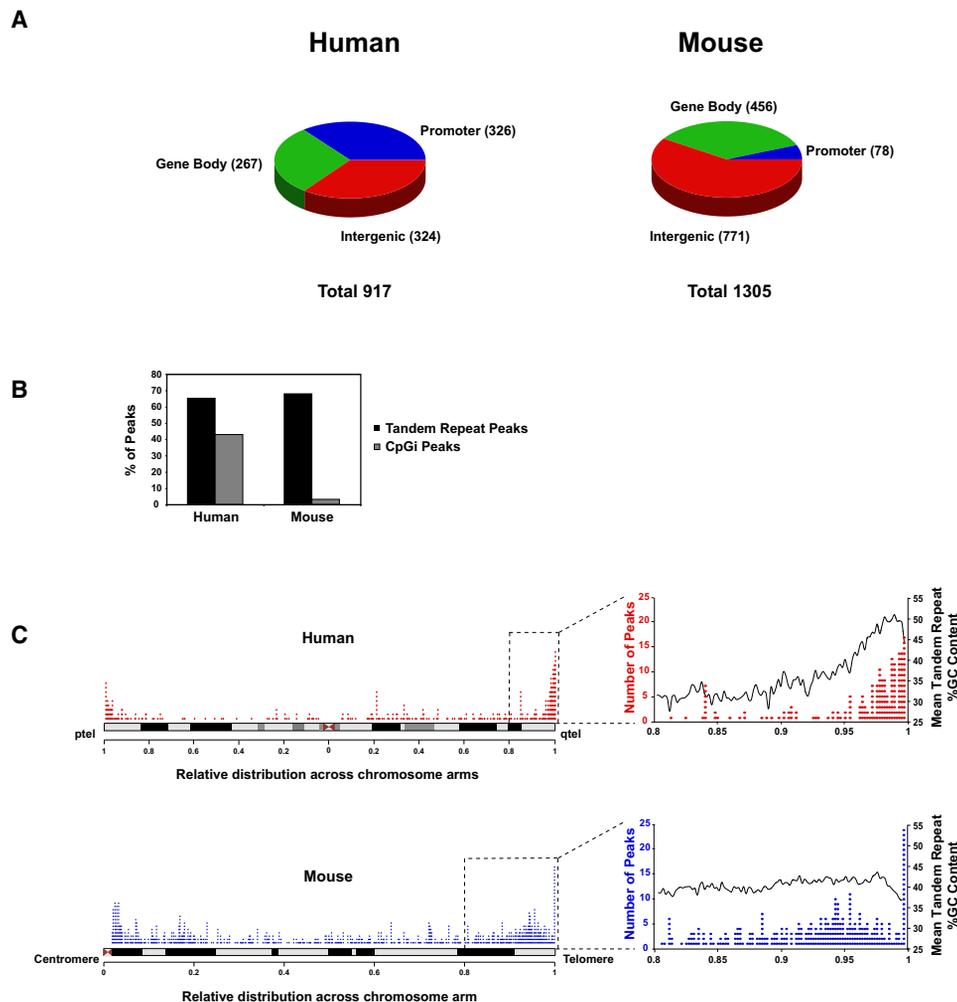
ChIP analysis was performed with an ATRX antibody that recognizes a C-terminal epitope only present in the full-length ATRX isoform (Figure 1A). Western blot was used to confirm that this antibody immunoprecipitates ATRX with detection using an independent antibody (Figure 1B). ATRX ChIP enrichment at rDNA was measured in primary erythroblasts and Hep3B cells (Figure 1C). Consistent with its ubiquitous expression profile, ATRX binds rDNA in both cell types tested. It was of interest that the maximal binding of ATRX occurs at the transcribed region of the locus that is very rich in G and CpG nucleotides. These observations confirm the specificity of the ATRX C-terminal antibody, validate the ChIP assay, and identify the ribosomal genes as direct ATRX targets.

### ATRX Binds G-Rich Telomeric and Subtelomeric Repetitive DNA

Having validated the ATRX ChIP protocol, we next addressed whether, in addition to rDNA, other putative targets (heterochromatic repeats) identified by indirect immunofluorescence were similarly bound by ATRX. To accomplish this, we took a ChIP-seq approach using Illumina high-throughput, short read sequencing to analyze primary human erythroid cells and mouse ESCs.

ATRX ChIP DNA from human primary erythroid cells was sequenced alongside sonicated input DNA as a control. The short read mapping protocol used for ChIP sequencing (see below) routinely discards nonunique genomic matches, precluding analysis of direct binding to repeat sequences. To overcome this, we interrogated the ATRX ChIP read library for perfect sequence matches to a variety of tandem and interspersed repeat sequences. As a negative control, we used ChIP-seq data for YY1 and SCL, transcription factors that have no known role at heterochromatic repeats. YY1 and SCL ChIP DNA both showed low enrichment of telomeric and nontelomeric satellite sequences (Figure 1D and Table S1 available online). ATRX ChIP DNA showed striking enrichments for the G-rich telomeric (TTAGGG)<sub>n</sub> repeats (~16-fold relative to input DNA) and telomere-associated repeats (~10-fold relative to input) (Figure 1D and Table S1). Similar results were obtained from the analysis of ChIP-seq data from mouse ESCs (Figure 1E and Table S1). Further confirmation of the specificity of the ATRX ChIP was demonstrated by showing that ATRX enrichment was abolished when ChIP was performed in mouse ESCs in which full-length ATRX was knocked out (Figure S1).

These data therefore show that previously described immunofluorescence studies reflect the binding of ATRX to telomeric and subtelomeric repeat sequences. The presence of ATRX at the subtelomeric TAR1 repeats is consistent with previous observations that DNA methylation at subtelomeric repeats is altered in patients with ATR-X syndrome (Gibbons et al., 2000).



**Figure 2. Genome-wide Comparison of Human and Mouse ATRX-Binding Site Characteristics**

(A) Pie charts show the location of human and mouse ATRX-binding sites relative to genes.

(B) The proportion of human and mouse ATRX peaks overlapping with the two most common classes of human ATRX-binding sites, TRs and CpG islands (CpGi). See also Figure S2C for genomic features associated with peaks.

(C) Ideograms showing the relative distribution of ATRX-binding sites across all human and mouse chromosomes. Each column represents the total number of ATRX peaks within nonoverlapping 1/500 divisions of all chromosome arms. The zoomed panels show the telomeric region, overlaid with the mean %G+C content of all tandem repeats throughout the same regions, for the respective human and mouse chromosomes. The sharp peak of subtelomeric targets in mouse represents clusters of (TTAGGG)<sub>n</sub> adjacent to the telomeres of a subset of mouse chromosomes. See also Figure S2E for the distribution of TRs and Refseq genes near telomeres.

See also Figure S2A for validation of targets by Q-PCR, Figure S2B for examples of ATRX-binding sites, Figure S2D for trinucleotide content of DNA sequence underlying peaks, Figure S3A for histone modifications associated with peaks, and Figure S3B for histone H3.3 distribution associated with ATRX peaks.

### Genome-wide Targets of ATRX Include CpG Islands and G-Rich Tandem Repeats

Having established that ATRX binds G-rich repetitive elements associated with rDNA, telomeres, and subtelomeric repeats, ATRX ChIP and input sequence reads were aligned to the genome if five or fewer matches were detected (allowing for three base-pair mismatches). Peak calling was performed on the ATRX ChIP-seq alignments using an input correction penalty to deplete peaks overlying enrichments of input reads. The input correction penalty effectively eradicated many peaks overlying DNA where there were differences in copy number between

the reference genome and the sequenced genome (e.g., at pericentromeric satellite DNA).

Using these criteria in primary human erythroid cells we identified 917 ATRX-binding sites genome-wide. The ChIP enrichment at 14 sites (chosen to represent the different classes of targets discussed below) was validated using Q-PCR. ATRX binding at most of these sites was enriched above background levels 10/14 (false discovery rate 4/14; Figure S2A). Of the 917 ATRX peaks called, approximately a third (324) were intergenic, a third were present at promoter regions (326), and a third were in the bodies of genes (267) (Figure 2A). All peaks were then

examined for overlap with annotated genomic sequence features (Figures S2B and S2C). Two striking observations arise from this analysis: first, irrespective of location relative to genes, human ATRX-binding sites commonly coincide with CpG islands (Figure 2B); second, the predominant sequence feature that ATRX binds in gene bodies and intergenic regions is tandem repetitive DNA (Figure 2B and Figure S2C). Analysis of ATRX binding in mouse ESCs (Figure 2A) identified a larger number of ATRX targets (1305) and showed a similar enrichment at TRs but less so at CpG islands (Figure 2B) (which occur much less frequently in the mouse genome) (Waterston et al., 2002).

As the tandem repetitive ATRX targets at rDNA and telomeres are G rich, we reasoned that this might be a common property of other ATRX-bound TRs. To test this we calculated the tri-nucleotide sequence content of ATRX-bound tandem repetitive targets. ATRX-bound TRs in both mouse and human are significantly enriched for G and C and CpG, and they are depleted in A- and T-containing trinucleotides relative to randomly selected control repeats (Figure S2D and data not shown).

These findings are consistent with the observation that in human cells, many ATRX-bound promoters are associated with CpG islands. Genome-wide analysis (in human) showed that there are no chromatin modifications consistently associated with binding of ATRX. Chromatin marks found at the promoter and intragenic and intergenic binding sites show the characteristic chromatin modifications associated with such features (Figure S3A). Together the data suggest that ATRX interacts predominantly with G and C and CpG-rich sequences contained within TRs and promoters.

### The Distribution of ATRX-Binding Sites Differs between Human and Mouse, Reflecting the Different Distributions of G-Rich Tandem Repeats

Initial analysis of the human ChIP-seq data suggested that ATRX targets may be clustered at subtelomeric regions of the genome (Figure 1D). This was confirmed when the proportions of ATRX-binding sites were plotted as a function of their distance from the nearest telomere (pooling data for all telomeres) (Figure 2C). However, it has previously been shown that in humans, GC content, CpG density, G-rich minisatellites, and gene density are all increased in subtelomeric regions of the genome, and this was confirmed here (Figure 2C and Figure S2F). In fact, the distribution of ATRX targets in humans appears largely to reflect the increase in GC content and G-rich TRs observed toward telomeres rather than increased gene or general TR density (Figure 2C and Figure S2E).

To explore this further, we compared the data from human with those from mouse, a species with less extremes of GC content and a different distribution of G-rich repeats (Waterston et al., 2002). In mouse, the GC content of TRs is not increased toward telomeres but is more evenly distributed across each chromosome (Figure 2C). Although the majority of mouse targets are associated with CpG islands or TRs (as in human), the mouse ATRX targets are less concentrated at telomeres (Figure 2C). This more even distribution of ATRX targets in mouse is consistent with the more even distribution of GC content and G-rich repeats in mouse compared to human (Figure 2C). These findings focus attention on the fact that ATRX appears to bind

many G-rich TRs in different chromosomal environments rather than genes within subtelomeric regions per se.

### Analysis of H3.3 Distribution in the Absence of ATRX

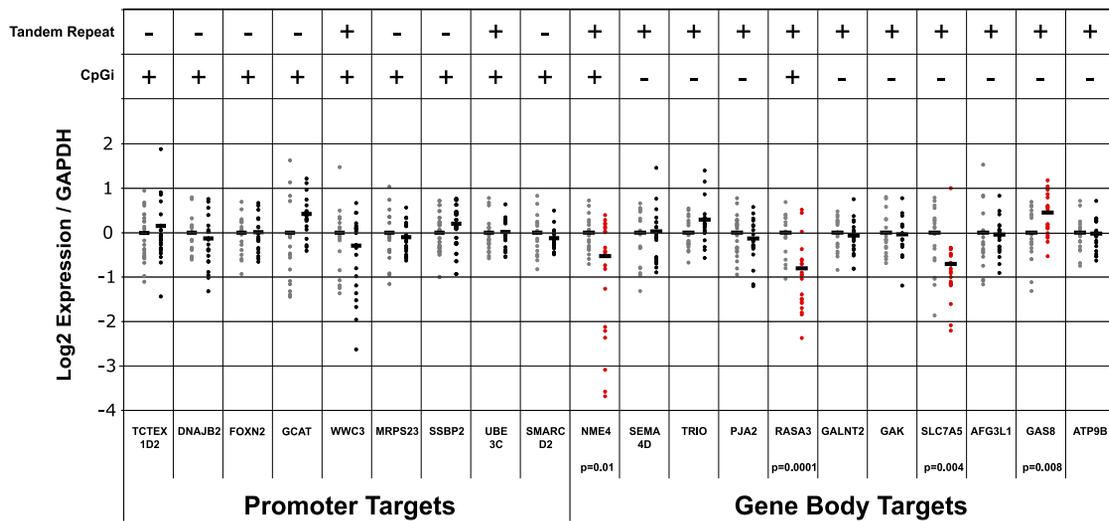
Telomeres are a site of rapid nucleosomal turnover as demonstrated by the incorporation of histone H3.3 (Goldberg et al., 2010). Furthermore, it has recently been shown that ATRX recruits the histone H3.3-specific chaperone DAXX and facilitates H3.3 deposition at telomeres and pericentric DNA (Drané et al., 2010; Lewis et al., 2010). In order to see whether H3.3 colocalized with ATRX at its target sequences (predominantly TRs, Figure 2B), data for the H3.3 distribution in mouse ESCs (Goldberg et al., 2010) were reanalyzed to determine the distribution of H3.3 at ATRX-binding sites (Figure S3B). Peaks of H3.3 are observed at genic and intergenic ATRX sites. ATRX has previously been shown to be required for H3.3 deposition at telomeres but not at promoters and transcription factor-binding sites (Goldberg et al., 2010). In order to see if the H3.3 distribution at these sites is dependent on ATRX, the patterns of H3.3 for *Atrx* and *Atrxnull* mouse ESCs were compared. The distribution of H3.3 is only subtly perturbed at ATRX-binding sites in gene bodies and intergenic sites (Figure S3B) with a slight diminution of the peak and increased signal in the adjacent sequence. If ATRX is required for H3.3 incorporation it may be only at a subset of these targets.

### Analysis of Expression of ATRX Targets when ATRX Is Mutated

Although we initially identified the human ATRX targets in erythroid cells, because many of the affected genes are widely expressed, we compared their expression in Epstein-Barr virus (EBV)-transformed lymphocytes from normal individuals ( $n = 19$ ) with expression in EBV cells from individuals harboring natural mutations in the *ATR-X* gene ( $n = 23$ ). Twenty ATRX targets (expressed in EBV-transformed lymphocytes) were chosen for analysis, including 9 ATRX promoter-binding targets and 11 tandem repetitive gene body targets. Four ATRX targets were significantly altered in expression in ATR-X patients relative to normal controls: *NME4*, *SLC7A5*, and *RASA3* were downregulated, whereas *GAS8* was upregulated (Figure 3). Interestingly all four novel targets contained tandem repetitive ATRX-binding sites, whereas none of the nonrepetitive, promoter-binding site target genes was affected. These data suggest that when ATRX alters gene expression, this involves an interaction with TRs associated with its target genes.

### ATR-X Exerts an Effect on Target Gene Expression via an Interaction with G-Rich Repeats

To examine the role of ATRX in regulating gene expression in detail, we analyzed the subtelomeric region of chromosome 16 (16p13.3), which contains two ATRX targets ( $\alpha$ -globin and *NME4*), both of which are downregulated in ATR-X syndrome. ChIP-seq analysis of this area was confirmed by ChIP-chip analysis (Figures 4A and 4B and Figure S4A). With this approach, three consistent peaks of ATRX binding were seen in primary erythroid cells. A small but reproducible enrichment was seen at the probe closest to the telomere (telomeric repeats were not included on the array). In erythroid cells, a broad region of



**Figure 3. Dysregulation of ATRX Targets Genes**

Q-PCR analysis of gene expression of ATRX ChIP target genes in ATR-X patient ( $n = 21$ ) and normal control ( $n = 19$ ) lymphoblastoid cDNAs. Gray dots represent control samples. Black dots represent genes unaffected in patient samples. Data are normalized to the mean values of the control samples. Black bars represent mean values. Red dots show genes affected in patient samples. p values are for a two-tailed Student's t test. The presence of a TR or CpG island underlying the ATRX-binding sites is indicated.

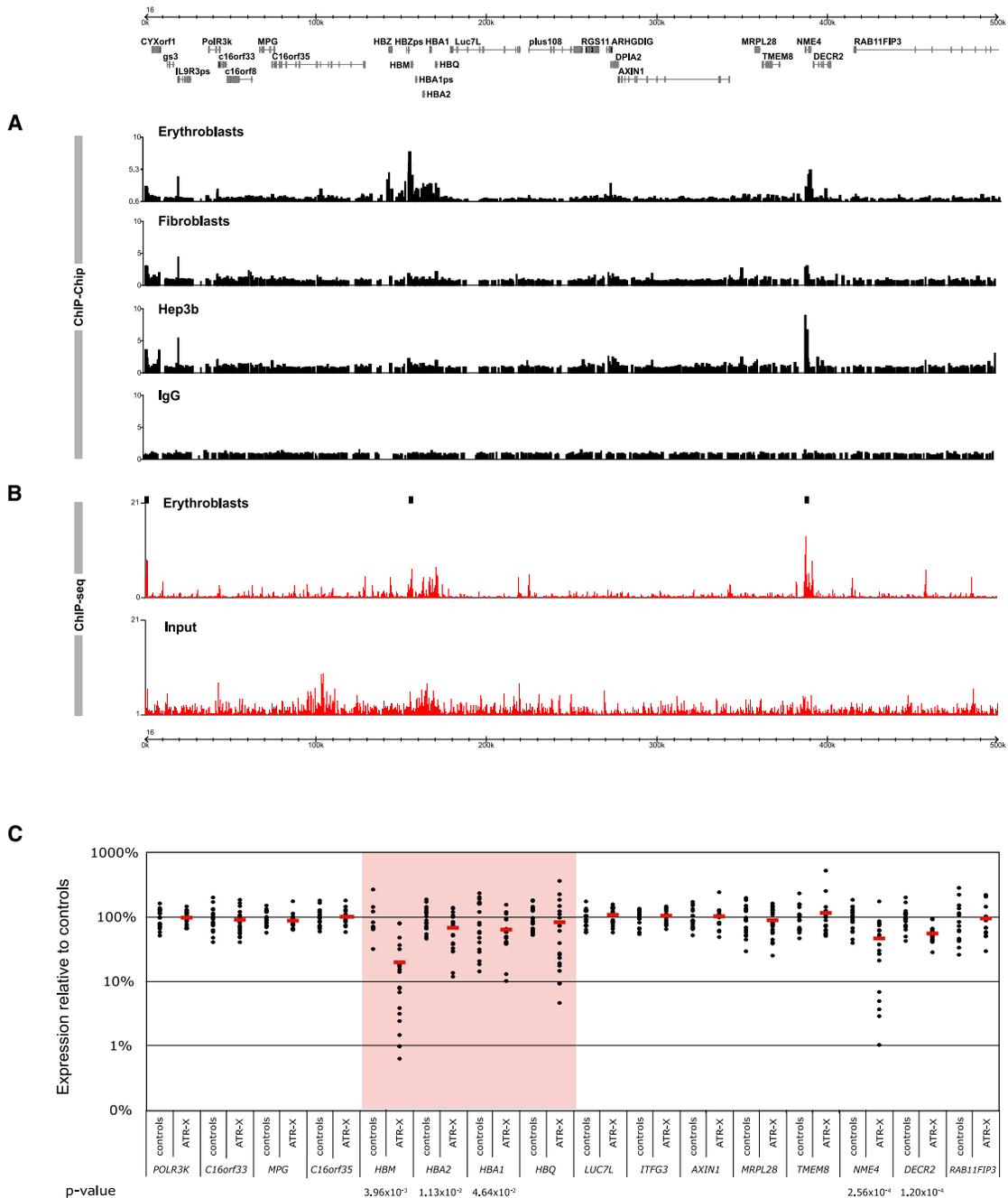
enrichment was seen across all the  $\alpha$ -like globin genes with maximum binding just upstream of the *HBM* globin gene. A third peak was seen at the gene encoding a nucleoside kinase, *NME4* (Figure 4A and Figure S4A). When we used Q-PCR (Figures S4B and S4C), we noted that all peaks of ATRX binding localized at or very close to regions of G-rich tandemly repetitive DNA. The sub-telomeric peak shows an enrichment lying  $\sim 150$  bp from the start of the telomeric satellite repeats (TTAGGG)<sub>n</sub> (Figure S4B). The maximum peak of binding in the  $\alpha$ -globin locus lies within a VNTR (CGCGGGGCGGGG)<sub>n</sub> 1 kb upstream from the *HBM* promoter, called  $\psi\zeta$  VNTR (Figures S4B and S4C). The peak at *NME4* is centered on an imperfect VNTR (CCCGGCCCCCA)<sub>n</sub> within the first intron of the gene (Figures S4B and S4C).

It has been previously shown that expression of RNA from the *HBA1* and *HBA2* globin genes is downregulated in patients with the ATR-X syndrome (Wilkie et al., 1990). However, maximal ATRX binding occurs not at the *HBA* genes but in close proximity to the *HBM* and *NME4* genes. We therefore took an unbiased approach using RT-PCR to measure expression of all 16 genes in the 500 kb region in normal individuals ( $n = 19$ ) and those proven to have ATR-X syndrome ( $n = 20$ ) (Figure 4C). Globin gene expression was analyzed using cDNA derived from erythroid cells, and other genes were analyzed using cDNA from EBV cell lines (nonglobin mRNAs are of very low abundance in erythrocytes). The most consistently and severely downregulated genes (*HBM* and *NME4*) were those associated with the greatest peaks of ATRX enrichment (Figures 4B and 4C). It was of interest that other significantly downregulated genes (*HBA2*, *HBA1*, *HBQ*, and *DECR2*) lie adjacent to these severely affected genes. Furthermore, the degree of downregulation of each  $\alpha$ -like globin (*HBM* > *HBA2* = *HBA1* > *HBQ*) gene is related to its proximity to the major peak of ATRX binding 1 kb upstream from the *HBM* gene.

This observation explains the  $\alpha$  thalassaemia seen in ATR-X syndrome and why  $\alpha$ -globin and not  $\beta$ -globin expression is perturbed, as only the former locus is associated with G-rich VNTRs (Higgs et al., 1998).

#### The Perturbation in Gene Expression Is Related to the Size of the Associated Tandem Repeat

In ATR-X syndrome,  $\alpha$ -globin RNA expression is often downregulated, but affected individuals show different degrees of repression (Figure 4C). This gives rise to different degrees of  $\alpha$  thalassaemia and is reflected by varying proportions of red cells containing HbH inclusions, ranging from 0%–30% (Gibbons et al., 2008). Importantly, such variation is seen between individuals with the same ATRX mutation (Figure S5A) and occurs both within and between affected families. However, for any individual, the level of HbH is relatively constant throughout life. If the downregulation of  $\alpha$ -globin expression in ATR-X syndrome resulted from a negative effect due to a TR then one might predict that the effect would be more extreme when the repeat is increased in size. The  $\psi\zeta$  VNTR is highly polymorphic. The size of the TR alleles was measured in 43 ATR-X individuals, and the average size in an individual was plotted against the level of HbH inclusions observed. A significant correlation ( $r$  value = 0.58;  $p = 0.0002$ ) was seen between the level of inclusions (reflecting the degree of  $\alpha$  thalassaemia) and the size of the TR (Figure 5A and Figure S5B).  $\psi\zeta$  VNTR lies within a block of linkage disequilibrium (Figure S5C and Table S3); polymorphisms within this block also show a correlation with the number of cells containing HbH inclusions. In contrast with  $\psi\zeta$  VNTR, another VNTR within this block, 3'HVR, showed a low correlation between size and the severity of  $\alpha$  thalassaemia (Figure S5B). Given the rapid evolution of VNTRs relative to the background haplotype, the strong correlation associated with the  $\psi\zeta$  VNTR strongly supports the

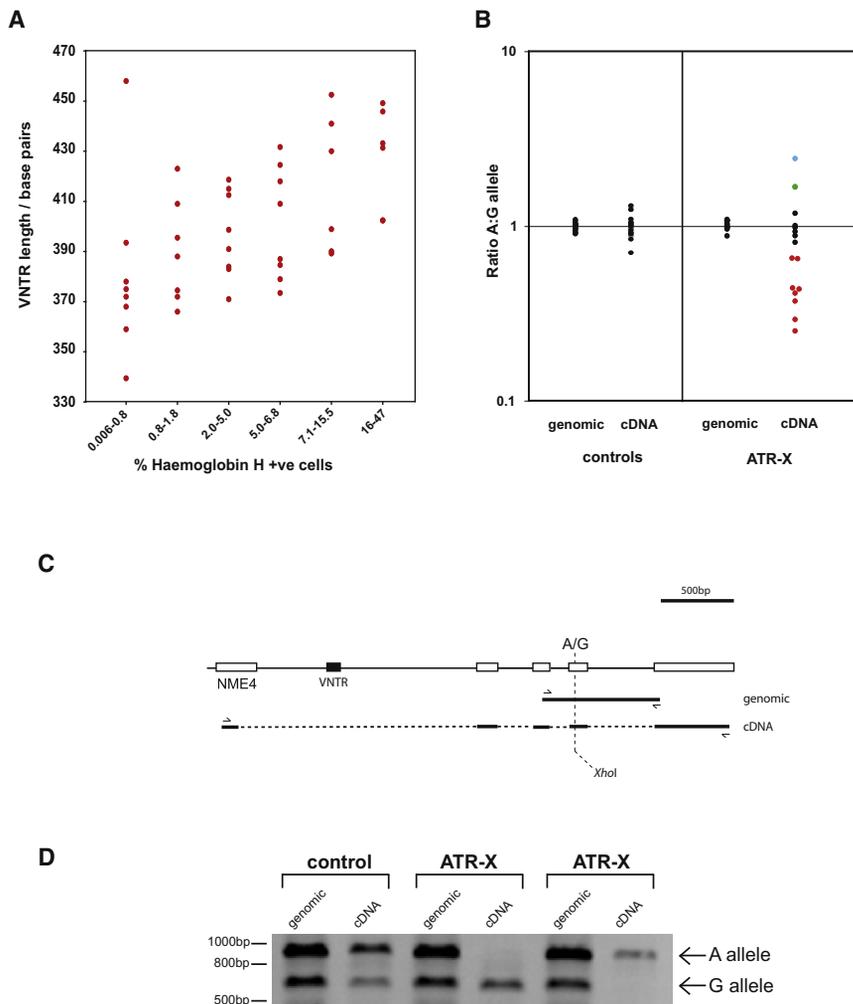


**Figure 4. ATRX Interacts with the  $\alpha$ -Globin Locus and Influences Gene Expression**

(A) Microarray analysis (black bars) of ATRX ChIP DNA enrichment across the 500 kb terminal region of chromosome 16p containing the  $\alpha$ -globin genes and surrounding ubiquitously expressed genes. ATRX ChIP DNA from erythroblasts (n = 4), fibroblasts (n = 1), and Hep3B (n = 2) cells were analyzed as well as erythroblasts immunoprecipitated with control IgG (n = 2). Representative data are shown. See Figure S4A for full dataset for erythroblasts.

(B) ChIP-seq analysis of erythroblast ATRX ChIP and input DNA using Illumina short-read sequencing. Graphs are a 50 bp sliding window of mapped reads. Black bars show peak calls.

(C) Q-PCR analysis of gene expression across the  $\alpha$ -globin gene locus in ATR-X patient (n = 20) and normal control (n = 19) cDNAs (from erythroid cells for the globin genes or lymphoblastoid cells for other genes). Expression was measured relative to GAPDH and the mean expression values for the normal controls were set to 100%. Red bars represent means of ATR-X patient expression and p values are for a two-tailed t test. See Figures S4B and S4C for validation of targets by Q-PCR and mapping of peaks to G-rich VNTRs.



**Figure 5. ATRX-Binding Variable Number Tandem Repeats Act as Length-Dependent Negative Regulators of Gene Expression When ATR-X Is Mutated**

(A)  $\psi\zeta$  VNTR length was measured in ATR-X patients with  $\alpha$  thalassaemia (n = 42) using PCR and agarose gel electrophoresis and plotted against the degree of  $\alpha$  thalassaemia as measured by % red cells showing Haemoglobin H inclusions. See Figure S5B to compare correlation of VNTR size and % red cells showing Haemoglobin H inclusions for  $\psi\zeta$  VNTR and 3'HVR. Spearman ranked correlation r value = 0.58, p value = 0.0002. See also Figure S5A for variable severity of  $\alpha$  thalassaemia in ATR-X syndrome, see Figure S5C and Table S3 for  $\alpha$ -globin locus haplotype and linkage analysis.

(B) Q-PCR-based allelic discrimination assay was used to determine the ratios of each *NME4* allele present in both genomic DNA and cDNA from controls and ATR-X patients. The y axis is the ratio of A:G allele (SNP rs14293), shown on a logarithmic scale. For control cDNA samples, the ratio of A:G allele expression is 0.70 to 1.30, mean = 1.0, n = 13. For ATR-X cDNA samples, the ratio of A:G allele expression is 0.24 to 2.37, mean = 0.84, n = 17. F-test p value =  $5.74 \times 10^{-5}$ . For the green datapoint, the larger VNTR is linked to the G allele; for the red datapoints, the larger VNTR is linked to the A allele. For the blue datapoint, alleles could not be discriminated based on VNTR size.

(C) Schematic representation of the exon/intron structure of *NME4*. White boxes represent exons. PCR amplicons are shown as generated from genomic DNA and cDNA. The presence of a polymorphic *XhoI* site generated by SNP rs14293 in *NME4* exon 4 is shown, which allows allelic discrimination by PCR amplification followed by an *XhoI* restriction digest assay, the restriction site being present in the G allele and abolished in the A allele. (D) Results show monoallelic expression of *NME4* in two individuals with ATR-X syndrome.

proposal that it is directly responsible for the variability seen in the level of HbH inclusions.

The effect of TR size was further examined at the *NME4* locus. Again the TR is highly polymorphic; in this case the presence of an expressed A/G single-nucleotide polymorphism (SNP) allowed us to determine the effect of the TR size on allele-specific expression. In ATR-X cases informative for the expressed SNP, the most downregulated allele is always in *cis* with the larger TR (Figures 5B and 5C). In some cases the expression was monoallelic (Figure 5D).

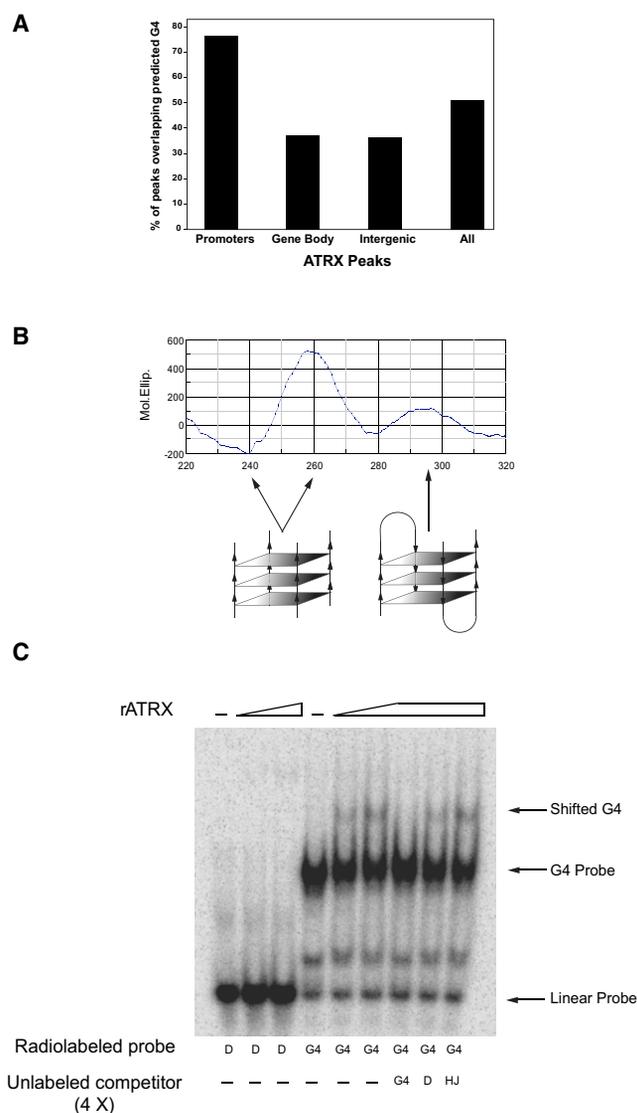
#### ATR-X Targets Have the Potential to Form G4 DNA, and ATRX Binds to G4 DNA Structures

Tandem repetitive sequences can take up a range of non-B DNA conformations (reviewed in Bacolla and Wells, 2009). G-rich sequences such as telomeres, rDNA, G-rich TRs, as well as CpG islands can form abnormal DNA structures *in vitro* referred to as G-quadruplex (G4) under physiological conditions (reviewed in Lipps and Rhodes, 2009). These structures form in G-rich

sequences that contain four tracts of at least three guanines separated by other bases and are stabilized by G-quartets that form between four DNA strands held together by Hoogsteen hydrogen bonds. Such structures are particularly likely to form when DNA becomes single stranded, for example during replication and transcription, and may interfere with these nuclear processes.

To explore the possibility that ATRX targets might form G4 structures *in vivo*, a genome-wide bioinformatic analysis using Quadparser was performed to identify regions that have the potential to form G4 DNA (Huppert and Balasubramanian, 2005). Fifty percent of ATRX peaks were found to overlap with putative quadruplex sequences (PQSs) (Figure 6A). Given the difficulty sequencing G-rich repeats and their consequent contraction in the reference genome, it is possible that PQSs are under-called in this analysis.

The potential for an ATRX-binding site to form G4 was further examined using circular dichroism (Paramasivan et al., 2007). The *NME4* TR is predicted to form G4. A 31 bp oligonucleotide



**Figure 6. ATRX Interaction with G-Quadruplex DNA**

(A) The proportion of human ATRX ChIP-seq peak coordinates overlapping with predicted G-quadruplex (G4) forming sequence.

(B) Circular dichroism. The presence of a positive ellipticity maximum at 260 nm and a negative ellipticity minimum at 240 nm suggests a predominantly parallel G4 form. The small positive ellipticity maximum at 295 nm is suggestive of the minor presence of an antiparallel G4 form. See Figure S6A for further examples of ATRX target sequences forming G4 structures.

(C) Gel-shift assay with recombinant full-length ATRX protein and a [ $\gamma$ - $^{32}$ P]ATP end-labeled G-rich oligonucleotide either preformed into a G4 structure (G4) or boiled and denatured (D). Reactions contained either 0, 2, or 4 nM rATR-X. Cold competition was performed with a 4-fold molar excess of either unlabeled G4 formed oligo (G4), denatured oligo (D), or a Holliday junction (HJ).

representing the repeat unit of the *NME4* TR was incubated in conditions that favor G4 DNA formation. The circular dichroism spectrum was obtained (Figure 6B). A positive ellipticity maximum was observed at 260 nm and a negative ellipticity minimum at 240 nm, consistent with a parallel G4 form. Another smaller ellipticity maximum at 295 nm suggested the coexis-

tence of an antiparallel G4 form. A further six ATRX TR target sequences were analyzed by circular dichroism (CD); the spectrographs were consistent with the formation of G4 including one sequence that was not predicted by Quadparser to form G4 (Figure S6A and Table S4).

Finally, we used a gel-shift assay to test whether ATRX could interact with G4 DNA in vitro. A G-rich oligonucleotide was preformed into G4 DNA, labeled, and incubated with full-length recombinant ATRX (Figure 6C). ATRX specifically bound to the G4 structure and no shift was observed when the structure was denatured by boiling before adding to the binding reaction. Further, binding to the formed G4 structure can be competed by a molar excess of unlabeled formed G4 but is less effectively competed by the denatured G4 oligonucleotide or another structured nucleic acid (Holliday junction) (Figure 6C), indicating that ATRX binds the G4 structure rather than the sequence per se. These data indicate that ATRX may be recruited to telomeres, other G-rich TR, and G-rich nonrepetitive DNA and interact with G-quadruplex DNA.

## DISCUSSION

Genome-wide analysis has shown that in euchromatin the predominant targets of ATRX are sequences containing VNTRs. Many of these are G and C rich with a high proportion of CpG dinucleotides. These observations explain why ATRX mutations affect the  $\alpha$ -globin cluster but not the  $\beta$ -globin cluster and cause  $\alpha$  thalassaemia. The  $\alpha$  cluster lies in a GC-rich subtelomeric region containing a high density of CpG islands and G-rich TRs that we have now shown are targeted by ATRX. The  $\beta$ -globin cluster has none of these features. It may also explain why in mouse there are a number of imprinted genes (that are also associated with tandemly repeated sequences) whose expression is affected by downregulation of ATRX (Kernohan et al., 2010).

The relationship between ATRX, VNTRs, and gene expression is clearly illustrated by the fact that of the targets whose expression was analyzed, all affected genes were associated with TRs. Furthermore, at some target genes, the degree by which gene expression is altered is directly related to the size of the VNTR, and in the case of one gene examined in detail (*NME4*), this can result in monoallelic expression. This provides an explanation for a long-standing question of why individuals with identical ATRX mutations have variable degrees of  $\alpha$  thalassaemia. As they all have the same mutation and apparently wild-type  $\alpha$ -globin gene clusters, one would have predicted that they would downregulate the  $\alpha$ -globin genes to the same extent. The highly significant relationship between the effect of the ATRX deficiency and the natural variation in the VNTR specifically explains the variable penetrance of ATR-X syndrome but more importantly identifies a new mechanism that might underlie many other genetic traits with similar variable penetrance.

A clearly demonstrated but unexplained phenomenon is that, in the absence of ATRX, expression of the target gene lying closest to an ATRX peak is the most severely perturbed. However, adjacent *cis*-linked genes (up to 10 kb downstream of the peak) are also affected. For example, although there is enrichment of ATRX across the entire  $\alpha$ -globin gene cluster, the main peak lies close to *HBM* and is associated with the G-rich TR in

the *HBZ* pseudogene. *HBM* is severely downregulated, but *HBA1* and *HBA2* are also downregulated to a lesser degree. Similarly, at *NME4*, although this gene is severely downregulated, the adjacent gene (*DECR2*) is also affected but to a lesser degree. It appears that ATRX normally binds to these G-rich TRs; in the absence of ATRX, the repeats at these loci now exert a repressive influence on transcription that spreads for some distance from the repeat.

At present it is not clear how ATRX might recognize such repeat sequences, but one possibility is that they form unusual, non-B DNA structures *in vivo*, and in the case of the G-rich repeats these may take the form of G-quadruplex structures. Such structures have been demonstrated *in vitro* using repeats from telomeres, rDNA, G-rich minisatellites, and CpG-rich promoters (all ATRX targets), and half of the ATRX targets identified here are predicted to form G4 DNA. In keeping with the observations described above, the longer the repeat the more likely it is to form G4 DNA (Ribeyre et al., 2009). Such structures have been notoriously difficult to identify *in vivo*, but the strongest evidence for their existence is at telomeres where it has been suggested that G4 structures may form during DNA replication and transcription (Lipps and Rhodes, 2009). It is therefore of interest that ATRX is recruited to telomeres during replication and that downregulation of ATRX by RNAi provokes a DNA-damage response (marked by gamma-H2AX) at telomeres during S phase (Wong et al., 2010). Downregulation of ATRX expression is also associated with an altered expression of telomere-associated RNA (Goldberg et al., 2010). Both of these observations would be consistent with ATRX playing a role in recognizing and/or modifying G4 structures at telomeres and by implication at other G-rich TRs *in vivo*. Nevertheless this is not the only factor determining the localization of ATRX, as at A/T-rich pericentric heterochromatin, the recruitment of ATRX depends on the presence of H3K9me3 (Kourmouli et al., 2005).

A role for ATRX at G-rich repeats may also be linked to the recent observation that ATRX is required for the incorporation of the histone variant H3.3 at telomeric repeats (Drané et al., 2010; Goldberg et al., 2010; Lewis et al., 2010). H3.3 may be incorporated into chromatin in a replication-independent or replication-dependent manner and has typically been found at actively transcribed regions of the genome and regions of inherent nucleosome instability where there is a rapid turnover of histones during interphase (Schneiderman et al., 2009). TRs with a propensity to form abnormal DNA structures are likely to be regions of rapid nucleosome turnover. An appealing hypothesis, therefore, is that ATRX influences gene expression by recognizing unusual DNA configurations at TRs and converting them to regular forms in part by facilitating incorporation of H3.3. Consistent with this we find that H3.3 is found at genic and intergenic ATRX-binding sites, the majority of which are TRs. However the distribution of H3.3 is only subtly perturbed at these sites when ATRX is disrupted. One possibility is that there is a critical requirement for ATRX at a subset of TRs (such as telomeres), whereas at other sites, other proteins can intervene. Future studies will focus on determining the role of ATRX in H3.3 deposition at specific sites.

The role of ATRX may be to recognize unusual forms of DNA and facilitate their resolution in several contexts. In the absence

of ATRX, G4 forms may persist and affect many nuclear processes including replication, transcription recombination, and repair.

## EXPERIMENTAL PROCEDURES

### Western Blotting

For ATRX western blotting, the mouse monoclonal 39c (McDowell et al., 1999) and rabbit polyclonal H-300 (Insight Biotechnology sc-1540) were used at 1:10 and 1:1000 dilutions, respectively. 23c and 39f recognize an epitope within ATRX and ATRXt N-terminal to the ADD domain, and H-300 recognizes a C-terminal epitope within 2193–2492 of full-length ATRX only.

### Immunopurification

Nuclear extracts were prepared from wild-type lymphoblastoid cells as previously described (Dignam, 1990) and incubated overnight at 4°C with H-300 antibody crosslinked to protein A-Sepharose. The beads were washed four times with 20 mM HEPES (pH 7.9), 0.5M KCl, 0.2 mM EDTA, 0.1% Tween, 0.5 mM DTT and immunoprecipitated protein eluted with 0.1 M glycine (pH 2.5), then neutralized with 1 M KHPO<sub>4</sub>. A mock immunopurification was performed as a control in the same way using normal rabbit IgG (Santa Cruz sc-2027) crosslinked to protein A.

### Chromatin Immunoprecipitation

ATRX chromatin immunoprecipitation was performed according to a published method (Lee et al., 2006) with the following modifications. Cells were fixed with 2 mM EGS (Pierce 26103) for 45 min at room temperature in PBS. Formaldehyde was then added to 1% for 20 min and quenched with 125 mM glycine. Chromatin was sonicated to under 500 bp and lysates were immunoprecipitated with 40 μg ATRX H300 (Insight Biotechnology sc-15408) antibody or rabbit IgG control (Dako X0903). DNA was precipitated with 20 μg of carrier glycogen and quantitated using a Qubit fluorimeter (Invitrogen).

### Real-Time Q-PCR

Real-time Q-PCR validation of ChIP-seq peaks was performed using SYBR green mastermix (Applied biosystems 4309155) or using Taqman probes with a 2× Taq mastermix (Applied Biosystems 4304437). SYBR green primers (Table S2) were designed using Macvector software and tested by running a five point, 8-fold serial dilution of genomic DNA to obtain a standard curve with  $r^2 > 0.99$ . PCR products were analyzed by melting curve and 3% agarose gel electrophoresis. Taqman probes were designed using Primer Express (Applied Biosystems). ChIP enrichments were determined relative to a 3 point dilution series of input DNA and normalized relative to GAPDH enrichment.

### Cell Culture

Human primary erythroblast cultures were prepared using a two-phase liquid culture system according to a published protocol (Fibach et al., 1991). HbH inclusions were detected in peripheral blood from ATR-X patients as previously described (Gibbons et al., 1992). Consent was obtained according to standard ethics approval guidelines.

### Microarray

Fluorescently labeled ChIP and input DNA was analyzed with a custom tiled microarray covering the subtelomeric region of human chromosome 16p as previously describe (De Gobbi et al., 2007).

### Gene Expression Analysis

RNA was extracted using Tri-reagent (Sigma) and quality checked by microfluidics separation using a 2100 Bioanalyser with an RNA 6000 nano kit (Agilent 5067-1511). One microgram was reverse transcribed with Superscript III (Invitrogen). Real-time RT-PCR was performed using commercial Taqman assays and custom assays. Primer sequences and product codes are listed in the Extended Experimental Procedures.

### High-Throughput Sequencing and Peak Analysis

See Extended Experimental Procedures.

### Allelic Discrimination

The ratio of allele-specific transcripts was ascertained with real-time technology, using an assay designed by Applied Biosystems (Table S2). In brief, a single amplicon was used, which in combination with two probes, each specific for one nucleotide of the polymorphism and labeled with a different fluorophore, allowed quantitation of each species. A standard curve with known ratios of A:G alleles was used to ensure specificity and quantitativeness of the assay, and results were confirmed with pyrosequencing (data not shown). Monoallelic expression is demonstrated with a restriction enzyme digest assay. The genomic PCR product is 846 bp, of which the G allele generates fragments of 581 bp and 265 bp when digested with XhoI. The cDNA PCR product is 854 bp, of which the G allele generates fragments of 563 bp and 291 bp when digested with XhoI. The A allele is undigested by XhoI in both cases.

### VNTR Size Measurement

$\psi\zeta$  VNTR allele lengths were measured in 43 ATR-X patients with  $\alpha$  thalassaemia by PCR and agarose gel electrophoresis. PCR was performed in 16.6 mM  $(\text{NH}_4)_2\text{SO}_4$ , 67 mM Tris-HCl (pH 8.8), 10% DMSO, 10 mM Beta mercaptoethanol, 125  $\mu\text{M}$  dNTP, 0.83 mM  $\text{MgCl}_2$ , 0.7 M Betaine, 0.3  $\mu\text{l}$  platinum Taq (Invitrogen), 250 nM primers 154505F/155293R (Table S2), and 100 to 400 ng genomic DNA in a 60  $\mu\text{l}$  reaction volume. 3' HVR allele sizes were measured by radio-labeled Southern blotting using AluI digested genomic DNA and a probe from  $\alpha\zeta$ 3'HVR.64 derived from genomic fragment Chr16:175999-177279. VNTR sizes were determined with a Typhoon 9400 Variable Mode Imager and ImageQuant TLv2005 software.

### Circular Dichroism Analysis

An oligonucleotide containing the repeat found within the VNTR of intron 1 of *NME4* (CCGGGGTGGGGGTGGGGGTGTGGGGGGTGA) was diluted to 2  $\mu\text{M}$  in 20 mM Tris HCl (pH 8) and 5 mM NaCl, heated to 95°C for 10 min then slowly cooled. CD analysis was performed as previously described using a Jasco 810 CD spectrometer (Giraldo et al., 1994).

### G4 Gel Shifts

G4 DNA was formed using oligonucleotide OX1-T (containing the *Oxytrichia* telomeric repeat sequence) and its structure confirmed as previously described (Sun et al., 1998). A Holliday junction structure was formed as previously described (Bachrati and Hickson, 2006). All DNA substrates were gel-purified prior to use. G4 DNA was labeled with [ $\gamma$ - $^{32}\text{P}$ ]ATP using T4 polynucleotide kinase, and unincorporated nucleotides were removed using a Sephadex G50 column. Where indicated the G4 probe was boiled for 10 min and quenched on ice to denature the G4 structure. Binding reactions (10  $\mu\text{l}$  volume) contained 2 fmol of  $^{32}\text{P}$ -labeled G4 DNA, full-length rATR-X protein as indicated (0, 20, or 40 fmol), 6 fmol  $T_{25}$  oligonucleotide to minimize nonspecific binding in a buffer containing 33 mM Tris acetate (pH 7.9), 66 mM Na acetate, 1 mM  $\text{MgCl}_2$ , 100  $\mu\text{g}/\text{ml}$  BSA, and 1 mM DTT. Where indicated, unlabeled competitor DNA (G4, denatured G4, or Holliday junction) was added to the reaction at 4-fold molar excess. Reactions were incubated on ice for 30 min. To each reaction 1  $\mu\text{l}$  of 50% glycerol was added and samples were loaded onto a 5% acrylamide gel and electrophoresed in 0.5  $\times$  TBE at 5 V/cm for 4 hr at 4°C. The gel was dried on Whatman filter paper and visualized by autoradiography.

### ACCESSION NUMBERS

Our ChIP-seq and microarray datasets have been deposited in the GEO database with accession number GSE22162.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and four tables and can be found with this article online at doi:10.1016/j.cell.2010.09.023.

### ACKNOWLEDGMENTS

We thank the members of the families studied for their participation; S. Butler and J. Sloane-Stanley for tissue culture; C. Wippo, J. Huddleston, and L. Marcelline for genotyping; P. Vathesatogkit and R. Totong for SNP validation; A. Goriely for assistance with pyrosequencing; C. Bachrati for gel-shift probes; and W. Wood for critical reading of the manuscript. In addition to others, the work was supported by the Medical Research Council and the National Institute of Health Biomedical Research Centre Programme. V.V. is supported by Thailand Research Fund (TRF) and BIOTEC, Thailand. K.M.L. was supported by an Oxford Nuffield Medical Fellowship, Oxford University.

Received: April 29, 2010

Revised: August 3, 2010

Accepted: September 13, 2010

Published: October 28, 2010

### REFERENCES

- Argentaro, A., Yang, J.C., Chapman, L., Kowalczyk, M.S., Gibbons, R.J., Higgs, D.R., Neuhaus, D., and Rhodes, D. (2007). Structural consequences of disease-causing mutations in the ATRX-DNMT3-DNMT3L (ADD) domain of the chromatin-associated protein ATRX. *Proc. Natl. Acad. Sci. USA* *104*, 11939–11944.
- Bachrati, C.Z., and Hickson, I.D. (2006). Analysis of the DNA unwinding activity of RecQ family helicases. *Methods Enzymol.* *409*, 86–100.
- Bacolla, A., and Wells, R.D. (2009). Non-B DNA conformations as determinants of mutagenesis and human disease. *Mol. Carcinog.* *48*, 273–285.
- Berube, N.G., Smeenk, C.A., and Picketts, D.J. (2000). Cell cycle-dependent phosphorylation of the ATRX protein correlates with changes in nuclear matrix and chromatin association. *Hum. Mol. Genet.* *9*, 539–547.
- Cui, K., Zang, C., Roh, T.Y., Schones, D.E., Childs, R.W., Peng, W., and Zhao, K. (2009). Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell* *4*, 80–93.
- De Gobbi, M., Anguita, E., Hughes, J., Sloane-Stanley, J.A., Sharpe, J.A., Koch, C.M., Dunham, I., Gibbons, R.J., Wood, W.G., and Higgs, D.R. (2007). Tissue-specific histone modification and transcription factor binding in alpha globin gene expression. *Blood* *110*, 4503–4510.
- Dignam, J.D. (1990). Preparation of extracts from higher eukaryotes. *Methods Enzymol.* *182*, 194–203.
- Drané, P., Ouararhni, K., Depaux, A., Shuaib, M., and Hamiche, A. (2010). The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. *Genes Dev.* *24*, 1253–1265.
- Fibach, E., Manor, D., Treves, A., and Rachmilewitz, E.A. (1991). Growth of human normal erythroid progenitors in liquid culture: a comparison with colony growth in semisolid culture. *Int. J. Cell Cloning* *9*, 57–64.
- Flaus, A., Martin, D.M., Barton, G.J., and Owen-Hughes, T. (2006). Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Res.* *34*, 2887–2905.
- Gibbons, R.J., Wilkie, A.O.M., Weatherall, D.J., and Higgs, D.R. (1991). A newly defined X linked mental retardation syndrome associated with a thalassaemia. *J. Med. Genet.* *28*, 729–733.
- Gibbons, R.J., Suthers, G.K., Wilkie, A.O.M., Buckle, V.J., and Higgs, D.R. (1992). X-linked a thalassaemia/mental retardation (ATR-X) syndrome: Localisation to Xq12-21.31 by X-inactivation and linkage analysis. *Am. J. Hum. Genet.* *51*, 1136–1149.
- Gibbons, R.J., Picketts, D.J., Villard, L., and Higgs, D.R. (1995). Mutations in a putative global transcriptional regulator cause X-linked mental retardation with a-thalassaemia (ATR-X syndrome). *Cell* *80*, 837–845.
- Gibbons, R.J., McDowell, T.L., Raman, S., O'Rourke, D.M., Garrick, D., Ayyub, H., and Higgs, D.R. (2000). Mutations in *ATR-X*, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. *Nat. Genet.* *24*, 368–371.

- Gibbons, R.J., Wada, T., Fisher, C., Malik, N., Mitson, M., Steensma, D., Goudie, D., Fryer, A., Krantz, I., and Traeger-Synodinos, J. (2008). Mutations in the chromatin associated protein ATRX. *Hum. Mutat.* *29*, 796–802.
- Giraldo, R., Suzuki, M., Chapman, L., and Rhodes, D. (1994). Promotion of parallel DNA quadruplexes by a yeast telomere binding protein: a circular dichroism study. *Proc. Natl. Acad. Sci. USA* *91*, 7658–7662.
- Goldberg, A.D., Banaszynski, L.A., Noh, K.M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X., et al. (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* *140*, 678–691.
- Higgs, D.R., Sharpe, J.A., and Wood, W.G. (1998). Understanding alpha globin gene expression: a step towards effective gene therapy. *Semin. Hematol.* *35*, 93–104.
- Huppert, J.L., and Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* *33*, 2908–2916.
- Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* *110*, 462–467.
- Kernohan, K.D., Jiang, Y., Tremblay, D.C., Bonvissuto, A.C., Eubanks, J.H., Mann, M.R., and Berube, N.G. (2010). ATRX partners with cohesin and MeCP2 and contributes to developmental silencing of imprinted genes in the brain. *Dev. Cell* *18*, 191–202.
- Kourmouli, N., Sun, Y., van der Sar, S., Singh, P.B., and Brown, J.P. (2005). Epigenetic regulation of mammalian pericentric heterochromatin in vivo by HP1. *Biochem. Biophys. Res. Commun.* *337*, 901–907.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* *10*, R25.
- Lee, T.I., Johnstone, S.E., and Young, R.A. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat. Protoc.* *1*, 729–748.
- Lewis, P.W., Elsaesser, S.J., Noh, K.M., Stadler, S.C., and Allis, C.D. (2010). Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proc. Natl. Acad. Sci. USA* *107*, 14075–14080.
- Lipps, H.J., and Rhodes, D. (2009). G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.* *19*, 414–422.
- McDowell, T.L., Gibbons, R.J., Sutherland, H., O'Rourke, D.M., Bickmore, W.A., Pombo, A., Turley, H., Gatter, K., Picketts, D.J., Buckle, V.J., et al. (1999). 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.
- Paramasivan, S., Rujan, I., and Bolton, P.H. (2007). Circular dichroism of quadruplex DNAs: applications to structure, cation effects and ligand binding. *Methods* *43*, 324–331.
- Rhead, B., Karolchik, D., Kuhn, R.M., Hinrichs, A.S., Zweig, A.S., Fujita, P.A., Diekhans, M., Smith, K.E., Rosenbloom, K.R., Raney, B.J., et al. (2009). The UCSC Genome Browser database: update 2010. *Nucleic Acids Res.* *38*, D613–D619.
- Ribeyre, C., Lopes, J., Boule, J.B., Piazza, A., Guedin, A., Zakian, V.A., Mergny, J.L., and Nicolas, A. (2009). The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo. *PLoS Genet.* *5*, e1000475.
- Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., et al. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* *4*, 651–657.
- Schneiderman, J.I., Sakai, A., Goldstein, S., and Ahmad, K. (2009). The XNP remodeler targets dynamic chromatin in *Drosophila*. *Proc. Natl. Acad. Sci. USA* *106*, 14472–14477.
- Sun, H., Karow, J.K., Hickson, I.D., and Maizels, N. (1998). The Bloom's syndrome helicase unwinds G4 DNA. *J. Biol. Chem.* *273*, 27587–27592.
- Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* *420*, 520–562.
- Wilkie, A.O.M., Zeitlin, H.C., Lindenbaum, R.H., Buckle, V.J., Fischel-Ghodsian, N., Chui, D.H.K., Gardner-Medwin, D., MacGillivray, M.H., Weatherall, D.J., and Higgs, D.R. (1990). Clinical features and molecular analysis of the a thalassemia/mental retardation syndromes. II. Cases without detectable abnormality of the a globin complex. *Am. J. Hum. Genet.* *46*, 1127–1140.
- Wong, L.H., McGhie, J.D., Sim, M., Anderson, M.A., Ahn, S., Hannan, R.D., George, A.J., Morgan, K.A., Mann, J.R., and Choo, K.H. (2010). ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. *Genome Res.* *20*, 351–360.
- Xue, Y., Gibbons, R., Yan, Z., Yang, D., McDowell, T.L., Sechi, S., Qin, J., Zhou, S., Higgs, D., and Wang, W. (2003). The ATRX syndrome protein forms a chromatin-remodeling complex with Daxx and localizes in promyelocytic leukemia nuclear bodies. *Proc. Natl. Acad. Sci. USA* *100*, 10635–10640.