DNA–protein interactions regulate critical biological processes. Identifying proteins that bind to specific, functional genomic loci is essential to understand the underlying regulatory mechanisms on a molecular level. Here we describe a co-binding-mediated protein profiling (CMPP) strategy to investigate the interactome of DNA G-quadruplexes (G4s) in native chromatin. CMPP involves cell-permeable, functionalized G4-ligand probes that bind endogenous G4s and subsequently crosslink to co-binding G4-interacting proteins in situ. We first showed the robustness of CMPP by proximity labelling of a G4 binding protein in vitro. Employing this approach in live cells, we then identified hundreds of putative G4-interacting proteins from various functional classes. Next, we confirmed a high G4-binding affinity and selectivity for several newly discovered G4 interactors in vitro, and we validated direct G4 interactions for a functionally important candidate in cellular chromatin using an independent approach. Our studies provide a chemical strategy to map protein interactions of specific nucleic acid features in living cells.
Results

Design of co-binding-mediated protein profiling. A small molecule that binds a variety of G4 DNA target structures in cells could be functionalized to allow mapping of G4-interacting proteins in their native environment with minimal perturbation (Fig. 1b). We based our probe design on pyridostatin (PDS), a highly G4-selective small-molecule ligand that has been widely used to target DNA and RNA G4s in cells. We previously showed that a PDS derivative and a protein can simultaneously bind a G4 in vitro, which makes a promising molecular scaffold to detect co-binding proteins.

We prepared two G4-ligand probes, photoPDS-1 (1) and photoPDS-2 (2) (Fig. 2a), by tethering PDS to a click alkyne handle and a photoreactive aliphatic diazirine group, which is small and has excellent chemical stability, photolabelling efficiency and low background binding. Probe 1 has a short, two-carbon linker and a two-unit polyethylene glycol longer linker (12 atoms) to enable probing proteins at different distances from the G4 binding site. In addition, we prepared a photoactivatable control 3 (Fig. 2a) that lacks a G4 binding moiety.

First, we assessed the binding affinity and selectivity of the probes towards G4 structures using an established fluorescence resonance energy transfer melting assay. Compared with the parent compound PDS, both 1 and 2 retained the capacity to bind and stabilize a panel of G4 oligonucleotides (G4 Kit1, G4 Myc and G4 Tel) and showed negligible stabilization of double-stranded DNA (dsDNA) (Fig. 2b and Extended Data Fig. 1a). Furthermore, fluorescence quench binding assays confirmed that 1 and 2 exhibit strong and selective binding to different G4 structures (Supplementary Table 2), such as G4 Myc with an apparent dissociation constants ($K_d$) of 197 ± 10 nM and 439 ± 36 nM, respectively (Fig. 2c), comparable to that of PDS binding ($K_d = 168 ± 8$ nM; Extended Data Fig. 1b). In contrast, 3 showed no apparent G4 binding (Fig. 2b and Extended Data Fig. 1a,b).

Photoproximity labelling of a G4 binding protein in vitro. As a proof of concept, we tested the probes using the G4-specific antibody BG4 in vitro (Fig. 2d). BG4 was incubated with a folded G4 Myc oligonucleotide that forms a well-characterized G4 structure, as well as incubation with non-G4 control oligonucleotides, such as a mutated single-stranded Myc (ss mutMyc) and a double-stranded Myc (ds Myc). The presence or absence of G4 formation was confirmed by circular dichroism spectroscopy (Extended Data Fig. 1c). Probes 1 and 2, as well as control 3, were then incubated with the pre-incubated BG4–oligonucleotides mixtures and photocross-linked at 365 nm. For each case, the probe was subsequently conjugated with tetramethylrhodamine-azide (TAMRA-azide) via the copper-catalysed azide–alkyne cycloaddition click reaction, and the protein–oligonucleotide–probe mixtures were each separated by denaturing sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) and then visualized by in-gel fluorescence scanning. We observed dose-dependent labelling of G4-Myc-bound BG4 by both probes 1 and 2 (Fig. 2e and Extended Data Fig. 1d), whereas negligible labelling was observed for control 3 (Fig. 2e). In addition, no labelling was observed in the presence of the control oligonucleotides ss mutMyc and ds Myc or in the absence of an oligonucleotide. This demonstrates for both probes 1 and 2 that crosslinking is made possible by co-binding to a G4 structure. In the case of BG4, labelling by probe 1 with the short linker, also suggests that the probe and BG4 co-bind to G4s in close proximity. The proof-of-concept paved the way for experiments to identify G4 binding proteins in cells.

Global profiling of DNA G4-interacting proteins in cells. We next employed our approach to identify G4-interacting proteins in human cells. Embryonic kidney HEK293T cells were treated with probes 1 and 2, and control 3 (20 μM), followed by photocrosslinking at 365 nm. The nuclear extract was conjugated with TAMRA-azide via the copper-catalysed azide–alkyne cycloaddition.
reaction, separated by SDS–PAGE and visualized by in-gel fluorescence scanning (Fig. 3a)11. We observed distinct bands over a range of concentrations for both probes 1 and 2 (Fig. 3b) and Extended Data Fig. 2a,b, which confirmed specific protein labelling as well as a good cell permeability and nuclear uptake, although probe 1 displayed a slightly higher efficiency. In addition, the probes did not show cell toxicity under the treatment conditions employed (Extended Data Fig. 2c).

Next, to identify the target proteins captured by G4-ligand probes, we employed a label-free, quantitative liquid chromatography (LC)–MS proteomics approach4. After photocrosslinking and extraction of the nuclear lysate, proteins were conjugated to biotin-azole and affinity purified on streptavidin beads, followed by on-bead digestion and quantitative LC–MS/MS analysis (Fig. 3a). Proteins that were detected in at least two out of four biological replicates and revealed that the identified candidates are implicated in various downstream processes (Fig. 3f). In particular, we observed a high enrichment among transcription termination factors, DNA repair factors, and epigenetic regulators.

Analysis of the annotated biological processes (Methods) revealed that the identified candidates are implicated in various different nuclear processes (Fig. 3f). In particular, we observed a large number of proteins involved in transcription, which is consistent with the emerging role of DNA G4s in transcription regulation24. Among the enriched proteins from diverse functional classes (Fig. 3g), we identified 19 of previously reported G4 interactors, such as hnRNP A112 and nucleolin13. Importantly, we identified numerous novel candidate G4 interactors, such as a master epigenetic regulator UHRF1, transcription termination factor TTF2, ATP-dependent RNA helicases (for example, DDX1 and DDX24) and pre-mRNA-spooling factor RBM22, that have been shown to have a direct association with chromatin14. Interestingly, we also identified several subunits of the chromatin remodelling complex SWI/SNF (SWItch/sucrose non-fermentable), such as SMARC4 and SMARCC1, which have only recently been linked to DNA G4s15,14.

Characterization of candidate proteins in vitro. Candidate G4-interacting proteins identified by co-binding-mediated proximity labelling could potentially bind to G4 directly or as part of a protein complex bound to G4 or in close proximity to G4s. To better characterize the binding properties for a selection of candidate proteins, we employed a selection of 3′–biotinylated, well-characterized G4 oligonucleotides that can form different types of G4 structures, which include parallel (Myc, Kit1 and Kit2), antiparallel (TBA) and hybrid (BCL2) G4s (Supplementary Table 3). The corresponding
mutated single-stranded mutant sequences that cannot fold into G4s and dsDNA were used as controls (Extended Data Fig. 3). The oligonucleotides were immobilized on streptavidin beads and used to affinity-enrich target proteins from HEK293T nuclear lysates, followed by western blot analysis. We investigated a selection of candidates identified by CMPP (SMARCA4, UHRF1, RBM22, TTF2, DDX24, DDX1 and HMGB2) that represent a variety of different functional protein classes (Fig. 3c,d). Strikingly, six out seven candidates showed G4-specific binding compared with that of the corresponding controls (Fig. 4a and Supplementary Table 4).
Fig. 4 | Validation of novel nuclear G4-selective binding proteins. **a.** Affinity enrichment coupled with western blot analysis of selected candidates for different topologies of G4 structures and control oligonucleotides (G-runs are highlighted in bold). A representative blot from two independent experiments with similar results is shown. **b–e.** Binding curves (the indicated \( K_d \) values were generated by ELISA) for the human recombinant full-length SMARCA4 protein to G4 Kit1, the single-stranded mutant (ss mutKit1) and double-stranded Kit1 (ds Kit1) (b), UHRF1 protein to G4 Kit1, ss mutKit1, Kit1 hemi-methylated dsDNA and ds Kit1 (c), DDX1 protein to G4 Myc, ss mutMyc and ds Myc (d), DDX24 protein to G4 Kit1, ss mutKit1 and ds Kit1 (e) and RBM22 protein to G4 NRAS and its mutant (mutNRAS) (f). Mean and error (± s.d.) are from three independent experiments (\( n = 3 \)). a.u., arbitrary units.

One protein, HMGB2, displayed single-stranded DNA and dsDNA, but no G4 binding (Extended Data Fig. 4a–c), which indicates that HMGB2 may bind to the dsDNA adjacent to G4s or to the single-stranded opposite strand. Intriguingly, all the other six G4 binding proteins displayed selectivity for different G4 topologies. Although SMARCA4, TTF2 and DDX24 each showed a preference for DDX1 are in line with its reported G4 binding affinity, which more strongly to all parallel G4s (Myc, Kit1 and Kit2) and well to hybrid-type G4 (BCL2) (Fig. 4a). Importantly, our findings for DDX1 are in line with its reported G4 binding affinity, which validates the approach45. Notably, RBM22 showed a particularly strong binding to G4 Kit1 with \( K_d = 1.2 \pm 0.2 \) nM, which is more than 7-fold lower than that of its known substrate hemi-methylated dsDNA \( (K_d = 8.5 \pm 1.1 \) nM) and 20-fold lower than its unmethylated duplex control \( (K_d = 21.2 \pm 3.5 \) nM) (Fig. 4c). Similarly, DDX1 and DDX2 showed a low nanomolar affinity to G4 Myc \( (K_d = 5.1 \pm 1.1 \) nM) and Kit1 \( (K_d = 58.2 \pm 14.1 \) nM), respectively (Fig. 4d,e). RBM22 selectively bound to both DNA and RNA G4s and a preference for RNA NRAS G4 \( (K_d = 52.1 \pm 11.3 \) nM) was observed (Fig. 4f and Extended Data Fig. 4d). Consistent with the affinity-enrichment experiments, considerably weaker or negligible binding was observed towards the control oligomers.

The affinity enrichment coupled with western blot analysis and ELISA experiments confirmed that our novel CMPP approach identifies genuine G4-interacting proteins in cells.

SMARCA4 binds at endogenous G4 in chromatin. Chromatin architecture is tightly linked to the presence of endogenous DNA G4s42 and may affect the binding of protein interactors. To further validate G4 binding interactions in a chromatin context, we focused on the candidate interactor SMARCA4, which is a part of the SWI/SNF chromatin remodelling complex that plays a key role in transcriptional regulation46. Given that endogenous G4s have recently
been mapped to open chromatin regions and promoters of highly expressed genes, suggesting that SMARCA4 may be linked to G4 function.

We focused on human K562 chronic myelogenous leukaemia cells in which we previously mapped endogenous G4s via G4 ChIP-seq\(^1\)–\(^10\). In this cell line, we performed SMARCA4 ChIP-seq and identified 28,265 SMARCA4 high-confidence binding sites from three biological replicates (Extended Data Fig. 5a). Strikingly, we observed that the majority of endogenous G4s (7,565 of 8,995, 84%) overlapped with SMARCA4 binding sites (Fig. 5a,b). Moreover, the SMARCA4 ChIP-seq signal was highly enriched and centred on endogenous G4 sites supportive of a direct SMARCA4-G4 binding interaction in chromatin (Fig. 5c). In contrast, no particular signal enrichment was observed at control sites that have the biophysical potential to form G4s (potential G4s, grey). The SMARCA4 ChIP-seq signal was highly enriched and centred on endogenous G4 sites (left) and potential G4s (right). (Fig. 5a,b). Moreover, the SMARCA4 ChIP-seq signal was highly enriched and centred on endogenous G4 sites supportive of a direct SMARCA4-G4 binding interaction in chromatin (Fig. 5c). In contrast, no particular signal enrichment was observed at control sites that have the biophysical potential to form G4s (potential G4s, grey). The SMARCA4 ChIP-seq signal was highly enriched and centred on endogenous G4 sites (left) and potential G4s (right).

Investigating SMARCA4 binding sites at different functional genomic regions, we observed the largest proportion of SMARCA4-G4 co-localization at promoters (42% of peaks), which suggests that these interactions may play a particular role in SMARCA4 promoter activity (Fig. 5d)\(^14\). In addition, although most SMARCA4 binding sites contained A/T-rich motifs (Extended Data Fig. 5b), a dominant G-rich motif was found in binding sites marked by endogenous G4s, which supports a direct binding to G4 structures and indicates an important alternative mode of recruitment to chromatin.

**Discussion**

Here we present a chemical CMPP approach to identify the cellular interactome of DNA G4 structures in native chromatin. The method employs functionalized, structure-specific small-molecule ligands that bind to G4s and mediate proximity labelling of endogenous G4 binding proteins via photactivatable diazirine groups. Compared with proteomic approaches carried out in vitro, the in situ capture in cells takes into account the local chromatin environment in a functioning cell and should also facilitate the detection of transient G4-protein interactions that are lost during cell lysis or washing steps.

Using the approach, we identified several hundred G4-associated proteins of which some were known G4-binding and many were not previously described. Several new G4 binding proteins were separately validated by in vitro assays and shown to be specific, high-affinity G4 binders. Given their distinct properties and various functions in biological processes, these proteins may play different key roles in regulation of the endogenous G4 landscape and G4 biology. The protein SMARCA4, which is part of a chromatin remodelling complex, was followed up further using genomic ChIP-seq methodology to demonstrate that SMARCA4 does, indeed, bind substantially to genomic sites in which G4 structures have been detected. This outcome confirms that our CMPP methodology does identify proteins that bind to G4 structures in cellular chromatin, particularly at gene promoters, and also implicates that SMARCA4-G4 interactions may be important for transcriptional control. Further experiments that involve protein knockdown or overexpression coupled with G4 ChIP-seq may ultimately help elucidate the associated mechanisms in more detail.

Although the CMPP probes were employed for relatively short treatment times, we cannot rule out the possibility that the ligands partially influence the endogenous G4 landscape and interactome in this study and in other work\(^{15}\), PDS and G4-interacting proteins have been shown to co-bind to the same G4 structure; however,

**Fig. 5** | **SMARCA4 is enriched at endogenous G4s.** a, Example genome browser view for XYLB, TMCC6 and LARP1. Signal tracks from ChIP-seq and control input as well as consensus peaks are shown for SMARCA4 (black) and G4s (blue). Sequences that have the biophysical potential to form G4s are shown for plus and minus strands (potential G4s, grey). In addition, although most SMARCA4 binding sites were separately validated by in vitro assays and shown to be specific, high-affinity G4 binders. Given their distinct properties and various functions in biological processes, these proteins may play different key roles in regulation of the endogenous G4 landscape and G4 biology. The protein SMARCA4, which is part of a chromatin remodelling complex, was followed up further using genomic ChIP-seq methodology to demonstrate that SMARCA4 does, indeed, bind substantially to genomic sites in which G4 structures have been detected. This outcome confirms that our CMPP methodology does identify proteins that bind to G4 structures in cellular chromatin, particularly at gene promoters, and also implicates that SMARCA4-G4 interactions may be important for transcriptional control.

Further experiments that involve protein knockdown or overexpression coupled with G4 ChIP-seq may ultimately help elucidate the associated mechanisms in more detail.
the situation can be more complex at high PDS concentrations, in which it has been shown to inhibit the binding of certain proteins to G4s. In addition, G4 ligands may induce the stabilization of weaker, more transient G4s or alter the folded topology of G4s in ways that may influence protein binding. For these reasons it is essential to validate candidate G4 interactors with orthogonal approaches in vitro and in untreated cells, as we show in this study. We were mindful of observations that prolonged treatment with G4 ligands can induce DNA damage and recruit associated proteins. Therefore, we limited ligand treatment times and concentrations to avoid potential artefacts and did not observe a particular enrichment of DNA damage-related proteins in our experiments.

In principle, the approach we describe here should be applicable to a wide range of cell types and cell states, which in turn may help reveal specific differences in G4 interactomes and biology. During the revision of this article, we became aware of an independent study that involved a pyrrolidine derivative of PDS and reported the identification of G4-related proteins in human SV589 and MM231 cells. Although we noted some overlap between the studies (61 shared protein candidates), which somewhat validates the independent approaches, most of the G4-associated proteins identified by our CMPP approach were not found in the independent study. The different outcomes may have arisen due to variations in protein expression levels, chromatin states and G4 biology between the different cell lines. There were also some important technical differences between the two studies, which may have contributed to differences in the outcomes. In our study, we fractionated the nuclear proteins to focus on chromatin-associated proteins involved in G4 biology, and also to minimize the masking of physiologically relevant DNA G4 interactors by high-abundance, cytosolic RNA-binding proteins (for example, ribosomal proteins and elongation factors). In addition, we employed the diazirine crosslinker control, which lacks a G4 binding moiety to account for and factor out background binding (Methods), as considerable off-target binding to diazirine photocrosslinkers has been reported previously.

Overall, our chemical method shows that it can provide an unbiased strategy for the global mapping of interacting proteins of nucleic acid structural features in live cells. Although this study focused on DNA G4 interactors, we also identified several candidates that are annotated as RNA-binding proteins. PDS can bind both DNA and RNA G4s with comparable affinity and, therefore, some of the identified proteins might, in principle, bind to nuclear RNA G4s. We envisage that future studies with RNA G4-specific probes might employ a similar approach to explore endogenous RNA G4-protein interactions. We also envision that the general principle will enable further studies to map endogenous interactomes of other nucleic acid structural features.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41557-021-00736-9.

Received: 20 January 2021; Accepted: 18 May 2021; Published online: 28 June 2021

References
1. Hudson, W. H. & Ortland, E. A. The structure, function and evolution of proteins that bind DNA and RNA. Nat. Rev. Mol. Cell Biol. 15, 749–760 (2014).
2. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome function and structure. Nature 357, 347–355 (2016).
3. Ummethum, H. & Hamperl, S. Proximity labeling techniques to study chromatin. Protot. Genet. 11, 450 (2020).
4. Mohammed, H. et al. Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. Nat. Protoc. 11, 316–326 (2016).
5. Raffée, M.-R., Girardot, C., Sigismundo, G. & Krijgsveld, J. Expanding the circuitry of pluripotency by selective isolation of chromatin-associated proteins. Mol. Cell 64, 624–635 (2016).
6. Engelken, E. et al. Proteins that bind regulatory regions identified by histone modification chromatin immunoprecipitations and mass spectrometry. Nat. Commun. 6, 7155 (2015).
7. Kim, D. I. & Roux, K. J. Filling the void: proximity-based labeling of proteins in living cells. Trends Cell Biol. 26, 804–817 (2016).
8. May, D. G., Scott, K. L., Campos, A. R. & Roux, K. J. Comparative application of BioID and TurboID for protein-proximity biotinylation. Cells 9, 1070 (2020).
9. Murale, D. P., Hong, S. C., Haque, M. M. & Lee, J. S. Photo-affinity labeling (PAL) in chemical proteomics: a handy tool to investigate protein–protein interactions (PPIs). Proteome Sci. 15, 1–34 (2017).
10. Nipakhis, M. J. & Cravatt, B. F. Enzyme inhibitor discovery by activity-based protein profiling. Annu. Rev. Biochem. 83, 341–377 (2014).
11. Ma, N. et al. Affinity-based protein profiling reveals cellular targets of phototoxic anticancer inhibitor SAPE. Chem. Commun. 14, 2546–2553 (2019).
12. Wang, Y. et al. Expedited mapping of the ligandable proteome using fully functionalized enantiomeric probe pairs. Nat. Chem. 11, 1113–1123 (2019).
13. Parker, C. G. et al. Ligand and target discovery by fragment-based screening in human cells. Cell 168, 527–541 (2017).
14. Sen, G. & Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. Nature 334, 364–366 (1988).
15. Spiegel, I., Adhikari, S. & Balasubramanian, S. The structure and function of DNA G-quadruplexes. Trends Chem. 2, 121–136 (2019).
16. Rodríguez, R. et al. Small-molecule-induced DNA damage identifies alternative DNA structures in human genomes. Nat. Chem. Biol. 14, 301–310 (2012).
17. Riffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. Nat. Chem. 5, 182–186 (2013).
18. Henderson, A. et al. Detection of G-quadruplex DNA in mammalian cells. Nucleic Acids Res. 42, 860–869 (2014).
19. Di Antonio, M. et al. Single-molecule visualization of DNA G-quadruplex formation in live cells. Nat. Chem. 12, 832–837 (2020).
20. Chambers, V. S. et al. High-throughput sequencing of DNA G-quadruplex structures in the human genome. Nat. Biotechnol. 33, 877–881 (2015).
21. Hänsel-Hertsch, R., Spiegel, J., Marsico, G., Tannahill, D. & Balasubramanian, S. Genome-wide mapping of endogenous G4 DNA structures by chromatin immunoprecipitation and high-throughput sequencing. Nat. Protoc. 13, 551–564 (2018).
22. Hänsel-Hertsch, R. et al. G-quadruplex structures mark human regulatory chromatin. Nat. Genet. 48, 1267–1272 (2016).
23. Hänsel-Hertsch, R. et al. Landscape of G-quadruplex DNA structural regions in breast cancer. Nat. Genet. 52, 878–883 (2020).
24. Varshney, D., Spiegel, I., Zynier, K., Tannahill, D. & Balasubramanian, S. The regulation and functions of DNA and RNA G-quadruplexes. Nat. Rev. Mol. Cell Biol. 1, 229–240 (2020).
25. Sun, H., Karow, J. K., Hickson, I. D. & Maiels, N. The Bloom’s syndrome helicase unwinds G4 DNA. J. Biol. Chem. 273, 27587–27592 (1998).
26. Fry, M. & Loeb, L. A. Human Werner syndrome DNA helicase unwinds G4 DNA. J. Biol. Chem. 274, 27587–27592 (1999).
27. Cogott, S. et al. MAZ-binding G4-decoy with locked nucleic acid and twisted intercalating nucleic acid modifications suppresses KRS in pancreatic cancer cells and delays tumor growth in mice. Nucleic Acids Res. 41, 4049–4064 (2013).
28. 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, 1999–1998 (2012).
29. Niu, K. et al. Identification of LARK as a novel and conserved G4 quadruplex binding protein in invertebrates and vertebrates. Nucleic Acids Res. 47, 7306–7320 (2019).
30. Mao, S. Q. et al. DNA G-quadruplex structures mold the DNA methylome. Nat. Struct. Mol. Biol. 25, 951–957 (2018).
31. Makowski, M. M. et al. Global profiling of protein–DNA and protein–nucleosome binding affinities using quantitative mass spectrometry. Nat. Commun. 9, 1653 (2018).
32. González, V., Guo, K., Hurley, L. & Sun, D. Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. J. Biol. Chem. 284, 23622–23639 (2009).
33. Williams, P. Li, L., Dong, X. & Wang, Y. Identification of SLIRP as a G-quadruplex-binding protein. J. Am. Chem. Soc. 139, 12424–12429 (2017).
34. Rodríguez, R. et al. A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. J. Am. Chem. Soc. 130, 15758–15759 (2008).
35. Yangyoru, P. M. et al. Dual binding of an antibody and a small molecule increases the stability of TERRA G-quadruplex. *Angew. Chem. Int. Ed.* **54**, 910–913 (2015).

36. Dubinsky, L., Krom, B. P. & Meijler, M. M. Diazirine based photoaffinity labeling. *Bioorganic Med. Chem.* **20**, 554–570 (2012).

37. Kleiner, P., Heydenreuter, W., Stahl, M., Korotkov, V. S. & Sieber, S. A. A whole proteome inventory of background photocrosslinker binding. *Angew. Chem. Int. Ed.* **56**, 1396–1401 (2017).

38. Mergny, J. L. & Maurizot, J. C. Fluorescence resonance energy transfer as a probe for G-quartet formation by a telomeric repeat. *ChemBioChem* **2**, 124–132 (2001).

39. Le, D. D., Di Antonio, M., Chan, L. K. M. & Balasubramanian, S. G-quadruplex ligands exhibit differential G-tetrad selectivity. *Chem. Commun.* **51**, 8048–8050 (2015).

40. Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise Huisgen cycloaddition process: copper(i)-catalyzed regioselective ‘ligation’ of azides and terminal alkynes. *Angew. Chem. Int. Ed.* **41**, 2596–2599 (2002).

41. Mishra, S. K., Tawani, A., Mishra, A. & Kumar, A. G4IPDB: a database for G-quadruplex structure forming nucleic acid interacting proteins. *Sci. Rep.* **6**, 38144 (2016).

42. Paramasivam, M. et al. Protein hnRNP A1 and its derivative Up1 unfold G-quadruplex DNA in the human KRAS promoter: implications for transcription. *Nucleic Acids Res.* **37**, 2841–2853 (2009).

43. Xiao, R. et al. Pervasive chromatin-RNA binding protein interactions enable RNA-based regulation of transcription. *Cell* **178**, 107–121 (2019).

44. Zyner, K. G. et al. Genetic interactions of G-quadruplexes in humans. *eLife* **8**, e46793 (2019).

45. Ribeiro de Almeida, C. et al. RNA helicase DDX1 converts RNA G-quadruplex structures into R-loops to promote IgH class switch recombination. *Mol. Cell* **70**, 650–662 (2018).

46. Clapier, C. R., Iwasa, J., Cairns, B. R. & Peterson, C. L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat. Rev. Mol. Cell Biol.* **18**, 407–422 (2017).

47. Marsico, G. et al. Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res.* **47**, 3862–3874 (2019).

48. Tolstorukov, M. Y. et al. SWI/SNF chromatin remodeling/tumor suppressor complex establishes nucleosome occupancy at target promoters. *Proc. Natl. Acad. Sci. USA* **110**, 10165–10170 (2013).

49. Di Antonio, M. et al. Selective RNA versus DNA G-quadruplex targeting by situ click chemistry. *Angew. Chem. Int. Ed.* **51**, 11073–11078 (2012).

50. Müller, S., Kumari, S., Rodriguez, R. & Balasubramanian, S. Small-molecule-mediated G-quadruplex isolation from human cells. *Nat. Chem.* **2**, 1095–1098 (2010).

51. Su, H. et al. Photoactive G-quadruplex ligand identifies multiple G-quadruplex-related proteins with extensive sequence tolerance in the cellular environment. *J. Am. Chem. Soc.* **143**, 1917–1923 (2021).

52. Mellacheruvu, D. et al. The CRAPome: a contaminant repository for affinity purification–mass spectrometry data. *Nat. Methods* **10**, 730–736 (2013).

53. Park, J., Koh, M., Koo, J. Y., Lee, S. & Park, S. B. Investigation of specific binding proteins to photoaffinity linkers for efficient deconvolution of target protein. *ACS Chem. Biol.* **11**, 44–52 (2016).

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) 2021
Co-binding-mediated proximity labelling of BG4. G4 Myc (7.3 μM) and the single-stranded mutated oligonucleotides were annealed in 10 mM Tris, pH 7.4, 200 mM KCl and ds Myc in 10 mM Tris, pH 7.4, 200 mM NaCl. The G4-specific antibody BG4 (5 μl of 6.6 μM in PBS) was then incubated with 5 μl of annealed oligonucleotides at room temperature by gently shaking for 1 h, followed by adding 5 μl of the indicated probes in 10 mM Tris Cl, pH 7.4, 100 mM KCl and incubated at room temperature for another hour. The solution was directly irradiated under 365 nm light for 10 min, and 1.7 μl of the ‘click’ mixture (2 μl of 50 mM CuSO4, in H2O, 2 μl of 50 mM TCEP in DMSO) and 5 μl of 2 mM TBT (tris-2-carboxyethylphosphine) in H2O, 1 μl of 10 mM TAMRA-azide in DMSO and 5 μl of 2 mM TBT (tris(1-benzyl-1H-1,2,3-triazolo-4-yl)methylamine) in 1/4 DMSO/t-BuOH) was added and the mixture was gently shaken at room temperature for 1 h. Next, 5 μl of 10% loading buffer (4x) was added and the solution was heated at 70°C for 10 min. Each sample (~2 μl) was loaded and separated by SDS-PAGE (NuPAGE 4 to 12% and Bis-Tris, 1.0 mm), visualized on a Bio-Rad ChemiDoc MP system and the obtained images processed using Image Lab (version 6.1.0) software. Three biological replicates were performed.

Proximity labelling of G4 interactions in live cells. The protocol was adapted from that described previously17. For gel-based experiments, HEK293T cells were grown in 6 cm dishes to ~90% confluence at the time of treatment. Cells were carefully washed with 5 ml of Dulbecco’s phosphate-buffered saline (DPBS) (GIBCO) and then incubated with the indicated probe-containing fresh FBS-free DMEM media (2.5 ml) at 37°C for 1 h, followed by direct irradiation under 365 nm light (UVPE-1000 Ultraviolet Crosslinker, Fisher Scientific) on ice for 10 min. To harvest cells in cold DPBS (3 ml) they were scraped, centrifuged (300g, 5 min, 4°C) and then washed with cold DPBS twice. Cell pellets were either treated directly or kept frozen at ~80°C until use. For MS-based experiments, a similar protocol as that above was used with minor modifications, which included that HEK293T cells were grown in 15 cm dishes to ~80% confluence and then treated with 15 cm fresh FBS-free media that contained the indicated probes.

Nuclear protein extraction for gel- and MS-based analysis. The cell pellets for 6 cm and 15 cm dishes were gently resuspended in 250 μl and 2.25 ml, respectively, of Hypotonic Buffer (10 mM HEPES, pH 7.4, 10 mM KCl and 1.5 mM MgCl2) with a protease inhibitor cocktail (PIC) (ThermoFisher, catalogue no. 78438) by pipetting several times and swelled on ice for 15 min. NP-40 (10%, 12.5 and 112.5 μl, respectively) was added and the pellets were vortexed at the highest setting for 10 s, centrifuged (900 g, 10 min, 4°C), and then washed with cold DPBS twice. Cell pellets were either treated directly or kept frozen at ~80°C until use. For MS-based experiments, a similar protocol as that above was used with minor modifications, which included that HEK293T cells were grown in 15 cm dishes to ~80% confluence and then treated with 15 cm fresh FBS-free media that contained the indicated probes.

Enrichment of probe-labelled nuclear G4 interactions for MS-based analysis. Nuclear proteins (700 μg) were diluted with 50 mM HEPES to 560 μl in a clean 5 ml microcentrifuge tube, to which 70 μl of 4% SDS 50 mM HEPES, pH 7.4, was added, followed by 10 μl of a freshly prepared click mixture (14 μl of 50 mM CuSO4 in H2O, 14 μl of 50 mM TCEP in H2O, 7 μl of 10 μl Biotin-PEG,–azide in DMSO and 35 μl of 2 mM TBT in 1/4 DMSO/t-BuOH). The mixture was incubated by rotating at room temperature for 1 h, followed by adding prechilled methanol (2.8 ml) and then left at ~20°C overnight for protein precipitation. The solution was centrifuged (16,000g, 10 min, 4°C) and the obtained protein pellets were washed with prechilled methanol (2.8 ml 2x). After drying at room temperature for 5 min, the nuclear proteins were redissolved in freshly prepared 0.2% SDS urea (625 μl, 6 M in DPBS) by sonication. The C-terminal peptides were then transferred to a 2 ml Protein LoBind tube, and the protein concentration was determined by a BCA protein assay.

Label-free quantitative proteomics data analysis. The label-free experiment consisted of 24 samples distributed in 6 groups, which included the treatments with the G4-ligand probes 1 and 2 and the negative control probe 3. Missing values for 3 are imputed by replacing them with the maximum value. The results were performed with volcano plots and Venn diagrams using the R libraries ggplot2 (https://cran.r-project.org/web/packages/ggplot2/index.html), ggrepel (https://cran.r-project.org/web/packages/ggrepel/index.html) and VennDiagram (https://cran.r-project.org/web/packages/VennDiagram/index.html). UniprotKB keywords of differentially expressed proteins were extracted using the Retrieve/ ID button on uniprot online functionality. The list of 79 G4-associated proteins in humans was downloaded from G4IPDB41 (accessed 20th November, 2020). The code is available on the github page dedicated to this study, https://github.com/slab-bioinformatics/cppm4

G4 affinity enrichment and western analysis. HEK293T cells were grown to ~70%–80% confluence at the time of treatment. Cell pellets of 10 million cells per 300 μl in a low salt buffer (20 mM HEPES, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.2 mM EDTA and 1 mM dithiothreitol) that contained PIC on ice for 15 min. Then, 15 μl of 10% NP-40 was added and pellets were vortexed for 1 min, centrifuged (900 g, 10 min, 4°C) to afford the nuclear pellets, which were then washed with low salt buffer. The nuclear pellets were lysed at a density of 30 million cells per 250 μl in high salt buffer (20 mM HEPES, pH 7.4, 500 mM NaCl, 3 mM MgCl2, 0.2 mM EDTA, 0.5% NP-40 and 1 mM DTT) that contained PIC by sonication in a Diagenode Bioruptor Plus (ten cycles, 30 s on and 30 s off at each high setting, 4°C). The lysates were centrifuged (16,000g, 10 min, 4°C) to afford the nuclear proteins, and the concentration was measured using the BCA protocol.

A slurry (50 μl) of Streptavidin MagPlex paramagnetic beads (Promega, catalogue no. Z5481) was prewashed with pull-down buffer (25 mM HEPES, 10.5 mM NaCl, 110 mM KCl, 1 mM MgCl2, 0.01 mM ZnCl2, 20% v/v glycerol, 0.1% Igepal C-630, 1 mM DTT and PIC) that contained 3% bovine serum albumin (BSA) and 0.2 g l−1 salmon sperm DNA (Invitrogen; catalogue no. 15632011) three times (2 ml), and then 75 μg of nuclear proteins was added into 500 μl of pull-down buffer that contained 3% BSA and 0.2 g l−1 salmon sperm DNA, and precleared by incubating with the prewashed beads at 4°C for 2h. Meanwhile, another 50 μl of beads was washed in the same manner as above. Then, 50 μl of 10 μg annealed biotinylated oligonucleotides (Sigma-Aldrich) was added to 500 μl of pull-down buffer preincubated with the rotation at room temperature for 30 min. The oligonucleotide immobilized beads were then washed with pull-down buffer (2 ml 5x) and incubated with the prewashed lysates (500 μl) by rotation at 4°C overnight. The beads were washed with cold pull-down buffer (500 μl 5x) and the biotinylated oligonucleotides on the beads were eluted in 25 μl
of LDS sample buffer that contained freshly prepared 50 mM DTT by heating at 70 °C for 10 min. Next, 3 μl of the LDS sample buffer were analysed with capillary electrophoresis in a Wes Simple Western system (ProteinSimple) according to the instructions of the manufacturer, or samples were kept frozen at −20 °C until analysis. The primary antibodies (Supplementary Table 4) and the corresponding secondary antibodies (anti-rabbit) were used to detect the target signal bands, which were analysed by the software Compass for SW (ProteinSimple).

**Enzyme-linked immunosorbent assay.** ELISAs for binding affinity and specificity were performed as described previously with minor modifications. Briefly, biotinylated oligonucleotides were bound to Pierce streptavidin-coated high capacity plates (ThermoFisher) followed by blocking with 3% BSA and incubation with full-length recombinant human GST-tagged UHRF1 (Abnova, catalogue no. H80029128-P01) and DDX24 (Abnova, catalogue no. H9057062-P01), His-tagged SMARCA4 (Abcam, catalogue no. ab82237), PBM22 (Origene, TP769856) and Myc/His-tagged DDX1 (Origene, TP308769) in ELISA buffer (100 mM KCl and 50 mM KH₂PO₄, pH 7.4). After three washes with the ELISA buffer, detection was achieved with an anti-GST HRP (horseradish peroxidase)-conjugated antibody (Abcam, catalogue no. ab4316) diluted to 1:5,000, anti-FLAG HRP-conjugated antibody (Abcam, ab2209, diluted to 1:100), and anti-HIS HRP-conjugated antibody (BioLegend, catalogue no. 652503) diluted to 1:3,000 in an ELISA buffer that contained 3% BSA and 3.3, 5.5-tetramethylbenzidine ELISA substrate (slow kinetic rate) (Abcam, ab171525). Signal intensity was measured at 450 nm on a SPECTROstar nano microplate reader (BMG Labtech). d values were calculated from binding curves assuming a one-site binding model in GraphPad Prism, and standard error of means from three replicates are reported.

**SMARCA4 ChIP-seq.** SMARCA4 ChIP-seq was performed essentially as described previously. Briefly, cells were first crosslinked in 2 mM disuccinimidyl suberate (ThermoFisher) for 10 min and then in 1% formaldehyde in the medium for 10 min at room temperature. The cells were quenched with 0.125 M glycine for 5 min and washed twice in ice-cold PBS. Chromatin was isolated and prepared using a ChIP-qPCR Kit (Chromatrap) and sonicated using a Bioruptor Plus (Diagenode) to an average DNA size of 150–400 base pairs. Magnetic protein G Dynabeads (ThermoFisher) were washed with PBS that contained 1% w/v BSA (Sigma-Aldrich), incubated with 5 μg of ChIP-grade antibody against SMARCA4 (Abcam, ab110641) for 1 h at room temperature and washed five times with PBS that contained 1% w/v BSA. Solubilized chromatin from 5 × 10⁴ cells was immunoprecipitated with antibody conjugated beads in RIPA buffer (50 mM Tris pH 7.4, 150 mM NaCl, 1% Igepal CA-630 and 0.5% sodium deoxycholate) for 12 h at 4 °C. Magnetic beads were washed 5× with RIPA buffer and chromatin was eluted. After crosslinking reversal, RNAase A (Ambion) and proteinase K (ThermoFisher) treatment, ChIP DNA was extracted using a Min-Elate purification kit (Qiagen). Sequencing libraries of ChIP DNA and input controls were generated using the NEBioNext Ultra DNA Library Prep Kit for Illumina (NE Biolabs) following the manufacturer’s protocol.

**SMARCA4 ChIP-seq data analysis.** Bioinformatics data analyses and processing were performed using Bash, R and Python programming languages. The following tools were used: cutadapt (version 1.10)²⁹, BWA (v0.7.15)¹, Picard (v2.14.0; https://broadinstitute.github.io/picard), MACS² (v2.1.1)¹, bedtools (v2.26.0), SAMtools (v1.3)¹, deepTools (v3.1.2)¹ and Intervene (v5.6.4)¹. Code is available in the github page dedicated to this study, https://github.com/sblab-bioinformatics/cmpp. Raw fastq files were trimmed with cutadapt to remove adapter sequences and low-quality reads (mapping quality <10). Reads were aligned to the human reference genome (version hg19) with BWA and duplicates marked using Picard (v2.14.0; http://broadinstitute.github.io/picard) and removed using SAMtools. G4 ChIP and SMARCA4 ChIP peaks were called by MACS² (q-value <0.05). Peak overlaps in different replicates were visualized with Intervene. Peaks were merged from replicates with bedtools and high confidence peaks were defined as those overlapping in two out of three replicates (SMARCA4) or five out of eight replicates (G4 ChIP-seq) as described previously. Fragment coverage bigWig files were computed at a 50 base pair resolution, 200 base pair average fragment size and normalization to sequencing depth (RPKM) using deepTools. Signal distribution from the SMARCA4 ChIP in K562 G4 ChIP-seq peaks and potential G4s was computed using the plotProfile function in deepTools.

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

**Data availability.** The label-free quantitative proteomics data reported in this study are included in Supplementary_Dataset_CMPP, which contains peptide intensities, metadata and enriched proteins from the 1 versus 3 and 2 versus 3 statistical comparisons. The SMARCA4 ChIP-seq data have been deposited in the NCBI GEO repository under accession number GSE165124. The BG4 ChIP-seq data were generated in a previous study and are available under accession number GSE107690. Source data are provided with this paper.

**Code availability.** For details about the bioinformatics data analysis, see https://github.com/sblab-bioinformatics/cmpp.
Extended Data Fig. 1 | Probes for co-binding-mediated proximity labelling of BG4 in vitro. a, Assessment of G4-ligand probes (1-3) of inducing thermal stabilization ($\Delta T_m$) on G4 Telo and G4 Myc using FRET melting assay. $\Delta T_m$ of 1 and 2 at 1 µM on G4 Telo are 25 °C and 27 °C, respectively. $\Delta T_m$ of 1 and 2 at 1 µM on G4 Myc, are 14 °C and 13 °C, respectively. While $\Delta T_m$ of 3 at 1 µM is 0. Mean is represented from two independent experiments (n = 2).

b, Assessment of G4-binding affinity of PDS and 3 using fluorescence titration binding assay by measuring apparent $K_d$ values. Mean and error (± S.D.) are represented from four independent experiments (n = 4).

c, Structure verification of G4 Myc, single-stranded mutMyc and double-stranded Myc with circular dichroism (G-runs are highlighted in bold). Mean of three independent experiments (n = 3) is represented.

d, Dose-dependent of CMPP of BG4 by 1 and 2. Signals from TAMRA and Coomassie staining represent probe-specific labelling and loading input, respectively. Representative images from three independent experiments with similar results are shown.
Extended Data Fig. 2 | Gel-based mapping of DNA G4-interacting proteins in human cells. a, Probe 1 and b, probe 2 display dose-depend protein labelling of nuclear proteomes in HEK293T cells. Representative gel images from three independent experiments with similar results are shown. c, CellTiter-Glo luminescent cell viability assay on probe treatment for 75 min to HEK293T cells under all conditions used in this study. Mean and error (± S.D.) are represented from four independent experiments (n = 4).
Extended Data Fig. 3 | Structure verification of oligonucleotides. CD spectra obtained here match previously reported spectra of the well-characterized DNA G4 sequences (G-runs are highlighted in bold, see Supplementary Table 3) with different topologies showing distinct bands65,66, including parallel a, G4 Myc, b, G4 Kit1 and c, G4 Kit2 by positive at ~260 nm and negative at ~240 nm; anti-parallel G4 TBA by positive at ~290 nm and ~240 nm, and negative at ~260 nm; d, hybrid G4 BCL2 by positive at ~290 nm and ~260 nm, and negative at ~240 nm. All G4 structures also share a positive band at ~210 nm. While the corresponding single-stranded mutant and duplex controls have lost these features. Mean of three independent experiments (n=3) is represented.
Extended Data Fig. 4 | Protein validation by affinity enrichment coupled with western blot analysis and ELISA. a, Affinity enrichment coupled with western blot analysis of HMGB2 for different topologies of G4 structures and control oligonucleotides. A representative blot from two independent experiments with similar results is shown. Structure verification of G4 Myc (b) and G4 Kit1 (c) and the indicated control oligonucleotides with CD spectroscopy. Curves are plotted by mean values of three independent experiments (n = 3). d, Binding curves with indicated dissociation constants (K_d) generated by ELISA for human recombinant full-length RBM22 protein to DNA G4 Myc, single-stranded mutant and Myc duplex DNA. Mean and error (± S.D.) are represented from three independent experiments (n = 3). G-runs are highlighted in bold.
**Extended Data Fig. 5 | Properties of SMARCA4 binding sites.**

**a.** Overlap of binding sites identified by SMARCA4 ChIP-seq in K562 chromatin across three biological replicates. Binding sites identified in at least two replicates were considered as high confidence binding sites. **b.** Binding motifs identified in SMARCA4 binding sites that are marked by or lack endogenous G4. The top3 motifs identified by EM for Motif Elicitation (MEME) analysis are shown.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a □ □

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided

*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested

- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) and variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

- For null hypothesis testing, the test statistic (e.g. f, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Software and code

Policy information about availability of computer code

Data collection

- Bruker 400 MHz Avance III HD Spectrometer, 500 MHz DCH Cryoprobe Spectrometer, Waters LCT Premier (ESI) spectrometer, Proteome Discoverer 2.2 (Thermo Scientific), Wes Protein Simple Western System with Compass for SW (4.0), Applied Photophysics Chirascan Plus Spectrometer, Bio-Rad CFX96 Touch Real-Time PCR Detection System, BMG PHERAstar Plus reader, BMG SPECTROstar nano microplate reader, Bio-Rad ChemiDoc MP system

Data analysis

- NMR data was processed in MestReNova (version 12.0.1)
- Gel analysis was performed with Image Lab (version 6.1.0)
- Binding data for FRET melting assay, fluorescence quench binding assay, ELISA as well as cell viability data were processed with Prism 7 (GraphPad Software Inc.).
- Capillary-based immunoassays on a Wes Protein Simple Western System were analysed using Compass for SW (4.0).
- Bioinformatics data analysis and processing were performed using Bash, R and Python programming languages. The following tools were also used: cutadapt (v 1.16), BWA (v 0.7.15), Picard (v 2.14.0), MACS2 (v 2.1.1), bedtools57 (v 2.26.0), samtools (V 1.6), deepTools (v 3.1.2) and intervene (0.6.4)
- Scripts for proteomics and CHP-seq analyses available in the github page dedicated to this study: https://github.com/silab-bioinformatics/cmpp

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data reported in this study are available at the NCBI GEO repository under accession number GSE165174. Results from the proteomics analysis are included as Supplementary Data Set.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/re-reporting-summary-faq.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | Initial in vitro validation of CMIP on B6 were performed in three independent replicates. Gel-based analyses of probe-labelled G4 interactomes were performed in three independent biological replicates. MS-based proteomics analyses of probe-labelled G4 interactomes were performed in four independent biological replicates. Initial studies screening of Protein-G4 interactions via pull-down affinity enrichment were performed two independent replicates. ELISAs were performed in three independent replicates to ensure reproducibility. ChIP-seq guidelines of the ENCODE consortium was followed. SMARCA4 ChIP-seq in K562 were performed in three biological replicates. All experiments provided consistent and reproducible measurements. |
| Data exclusions | No data were excluded from analyses. |
| Replication | All experimental findings were reliably reproduced as described above. |
| Randomization | No randomization were applied. |
| Blinding | No animals or human participants were used in studies, and blinding was not used. |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | n/a | Involved in the study |
| --- | --- | --- |
| Antibodies | ❌ | ❌ |
| Eukaryotic cell lines | ☑ | ☑ |
| Palaeontology and archaeology | ❌ | ☑ |
| Animals and other organisms | ❌ | ☑ |
| Human research participants | ☑ | ☑ |
| Clinical data | ☑ | ☑ |
| Dual use research of concern | ☑ | ☑ |

| Methods | n/a | Involved in the study |
| --- | --- | --- |
| ChIP-seq | ☑ | ☑ |
| Flow cytometry | ❌ | ☑ |
| MRI-based neuroimaging | ❌ | ☑ |

Antibodies

Antibodies used

Rabbit TIF2 polyclonal antibody (Proteintech, cat.#32103-1-AP), rabbit RBM23 polyclonal antibody (Proteintech, cat.#21033-1-AP), rabbit HMGB2 polyclonal antibody (Proteintech, cat.#18597-1-AP), rabbit monodonal HIF1 antibody (abcam, cat.#ab110641), rabbit monodonal UHRF1 antibody (abcam, cat.#ab194238), rabbit polyclonal DDX1 antibody (abcam, cat.#ab70252), rabbit polyclonal DDX24 antibody (abcam, cat.#ab70462), rabbit anti-GST HRP-conjugated antibody (abcam, cat.#ab3416), goat anti-FLAG HRP-conjugated antibody (abcam, cat.#ab1238), mouse anti-HIS HRP-conjugated antibody (BioLegend, cat.#652503), B64.
Eukaryotic cell lines

Policy information about: cell lines

Cell line source(s)  Human embryonic kidney HEK293T cells (ATCC, CRL-3216) were provided by the CRUK Cambridge Institute Biorepository Core Facility. Human chronic myelogenous leukemia K562 cells (CCL-243) were purchased from ATCC.

Authentication  Short tandem repeat (STR) profiling was used to distinguish between individual human cell lines and rule out intra-species contamination performed by the CRUK Cambridge Institute Biorepository Core Facility.

Mycoplasma contamination  Cells were tested mycoplasma-free based on RNA-capture ELISA performed by the CRUK Cambridge Institute Biorepository Core Facility.

Commonly misidentified lines (See INCLAC register)  No commonly misidentified cell lines were used.

ChIP-seq

Data deposition

☑ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☑ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Files in database submission

The data reported in this paper are available at the NCBI GEO repository under accession number GSE165124. The BG4 ChIP-seq data were generated in a previous study and are available under accession number GSE107690.

K562_Rep1/Input_SLX-19726.NEB/708_NEB/508.HLVFvB6XG.s_1.r_1.fq.gz
K562_Rep1_SMARCA4_SLX-19726.NEB/709_NEB/502.HLVFvB6XG.s_1.r_1.fq.gz
K562_Rep2/Input_SLX-19727.NEB/701_NEB/501.HNLSG06XG.s_1.r_1.fq.gz
K562_Rep2_SMARCA4_SLX-19727.NEB/707_NEB/507.HNLSG06XG.s_1.r_1.fq.gz
K562_Rep3/Input_SLX-19727.NEB/702_NEB/502.HNLSG06XG.s_1.r_1.fq.gz
K562_Rep3_SMARCA4_SLX-19727.NEB/708_NEB/508.HNLSG06XG.s_1.r_1.fq.gz
K562_SMARCA4_rep3-3_mult2003.bed

Genome browser session

Provide a link to an anonymized genome browser session for "initial submission" and "revised version" documents only, to enable peer review. Write "no longer applicable" for "final submission" documents.

Methodology

Replicates  Shared peaks for SMARCA4 ChIP-seq replicates are shown in Supplementary Fig. 5a.

Sequencing depth  ChIP-seq was performed in single-end mode with 75-bp read length. In accordance with ENCODE guidelines for TF and narrow-peak histone ChIP-seq experiments, we aimed for >10 Mio usable fragments per library.

K562_Rep1/Input_SLX-19726.NEB/708_NEB/508.bam 42,432,209
K562_Rep1_SMARCA4_SLX-19726.NEB/709_NEB/502.bam 46,790,214
K562_Rep2/Input_SLX-19727.NEB/701_NEB/501.bam 31,733,009
K562_Rep2_SMARCA4_SLX-19727.NEB/707_NEB/507.bam 49,046,604
K562_Rep3/Input_SLX-19727.NEB/702_NEB/502.bam 80,095,823
K562_Rep3_SMARCA4_SLX-19727.NEB/708_NEB/508.bam 41,693,339

Antibodies  Per replicate we employed 5 μg of ChIP-grade antibody against SMARCA4 (abcam, ab110641)

Peak calling parameters  macs2 callpeak --name ./macs2_output/$%n%nodup.bam nodup.cq005.all -t $f cp input.nodup.bam --format BAM --size 'hs' -- bw=300 --qvalue 0.05

Data quality  (see also Supplementary Fig. 5a)

Individual peak numbers for qvalue <0.05:
21,487 Rep1_SMARCA4
27,356 Rep2_SMARCA4
40,707 Rep3_SMARCA4

Software  The following tools were used for ChIP-seq analysis: cutadapt (v 1.18), BWA (v 0.7.15), Picard (v 2.14.0), MACS2 (v 2.1.1), bedtools (v 2.26.0), samtools (v 1.6) and deepTools (v 3.1.2).