

G-quadruplex structures mark human regulatory chromatin

Robert Hänsel-Hertsch¹, Dario Beraldi¹, Stefanie V Lensing¹, Giovanni Marsico¹, Katherine Zyner¹, Aled Parry¹, Marco Di Antonio², Jeremy Pike¹, Hiroshi Kimura³, Masashi Narita¹, David Tannahill¹ & Shankar Balasubramanian^{1,2,4}

G-quadruplex (G4) structural motifs have been linked to transcription^{1,2}, replication³ and genome instability^{4,5} and are implicated in cancer and other diseases^{6–8}. However, it is crucial to demonstrate the bona fide formation of G4 structures within an endogenous chromatin context^{9,10}. Herein we address this through the development of G4 ChIP-seq, an antibody-based G4 chromatin immunoprecipitation and high-throughput sequencing approach. We find ~10,000 G4 structures in human chromatin, predominantly in regulatory, nucleosome-depleted regions. G4 structures are enriched in the promoters and 5' UTRs of highly transcribed genes, particularly in genes related to cancer and in somatic copy number amplifications, such as *MYC*. Strikingly, *de novo* and enhanced G4 formation are associated with increased transcriptional activity, as shown by HDAC inhibitor-induced chromatin relaxation and observed in immortalized as compared to normal cellular states. Our findings show that regulatory, nucleosome-depleted chromatin and elevated transcription shape the endogenous human G4 DNA landscape.

Regulatory regions in chromatin are characterized by nucleosome depletion that allows access of proteins directing gene transcription, replication and epigenetic plasticity¹¹. These accessible regions are quantitative indicators of cellular fate, origin and identity¹², yet the underlying DNA structural features remain largely uncharacterized. G4 DNA structures have recently been visualized in human cells^{13,14} and directly mapped in purified genomic DNA¹⁵. Employing G4-promoting conditions, ~700,000 G4-induced polymerase-stalling sequences have been observed in the human genome *in vitro* using purified single-stranded DNA as a template for G4-seq, a high-throughput sequencing method for use in the discovery of structural features in DNA¹⁶. G4 structures have been computationally predicted to form in replication origins^{17,18} and nucleosome-depleted regions¹⁹. Proteins such as human *ATRX*²⁰, *XPB* and *XPB*¹ and yeast *Pif1* (ref. 21) and *Rif1* (ref. 22) have been mapped by ChIP-seq to G-rich genomic sequences predicted to adopt G4 structures *in vivo*. However, direct capture of G4 structure formation at high resolution within a chromatin context is missing.

We set out to elucidate the relationship between G4 structure, chromatin and transcription in human cells. To achieve this, we developed

a G4 ChIP-seq protocol employing an engineered antibody specific to G4 structures (BG4)¹³ to map the genome-wide location of G4 structures in the chromatin of the spontaneously immortalized, non-oncogenic human epidermal keratinocyte HaCaT cell line²³ (Fig. 1 and Supplementary Fig. 1). We mapped nucleosome-depleted regions, using formaldehyde-assisted isolation of regulatory elements with sequencing (FAIRE-seq)²⁴ and the assay for transposase-accessible chromatin with sequencing (ATAC-seq)²⁵ (Supplementary Fig. 1), and determined transcriptional output by RNA-seq. Overall, in HaCaT chromatin, 10,560 high-confidence G4 ChIP-seq peaks were identified, with 87% conforming to a motif that forms G4 structure *in vitro* (Fig. 1 and Supplementary Fig. 2; see Online Methods, 'G4 motif and enrichment analysis')¹⁶. Analysis of G4 motifs showed that 21% of all G4 ChIP-seq peaks contained G4 motifs with loop sizes of 1–7 bases (i.e., canonical G4 motifs; G₃N_{1–7}G₃N_{1–7}G₃N_{1–7}G₃) (Supplementary Fig. 2)²⁶. Canonical G4 motifs were more enriched in G4 ChIP-seq peaks than were other G4 motifs (Supplementary Fig. 2), such as those with longer loops²⁷ or bulges in G tracts²⁸. We also validated the G4 structure-specific enrichment observed in the HaCaT G4 ChIP-seq peaks in control experiments with G4 ChIP-qPCR, including ones where BG4 was preincubated with a G4-forming or single-stranded DNA sequence before G4 ChIP (Supplementary Fig. 3), and by bioinformatic analysis of recurrent MEME-derived motifs and their density in identified G4 ChIP-seq peaks (Supplementary Fig. 2). Strikingly, most (98%) G4 ChIP-seq peaks overlapped with regions that were identified by either FAIRE-seq or ATAC-seq, or both (Fig. 1b). We considered the possibility that BG4 might simply target the most accessible open chromatin regions, rather than only true G4 sites. To rule this out, we overlapped the high-confidence G4 ChIP-seq peak set (10,560) with the 10,560 most accessible FAIRE sites. FAIRE accessibility was ranked according to *q* value (minimum false discovery rate, FDR) for peak enrichment assigned by the MACS2 peak caller. We found that ~44% of the G4 ChIP-seq peaks did not overlap with these highly accessible FAIRE sites, but were instead found among relatively less accessible FAIRE sites rather than the 10,560 most accessible FAIRE sites. This suggests that chromatin accessibility alone is not sufficient for BG4 binding. We further validated the presence of G4 structures in nucleosome-depleted chromatin using immunofluorescence

¹Cancer Research UK, Cambridge Institute, Li Ka Shing Centre, Cambridge, UK. ²Department of Chemistry, University of Cambridge, Cambridge, UK. ³Cell Biology Unit, Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, Japan. ⁴School of Clinical Medicine, University of Cambridge, Cambridge, UK. Correspondence should be addressed to S.B. (sb10031@cam.ac.uk).

Received 9 June; accepted 8 August; published online 12 September 2016; doi:10.1038/ng.3662

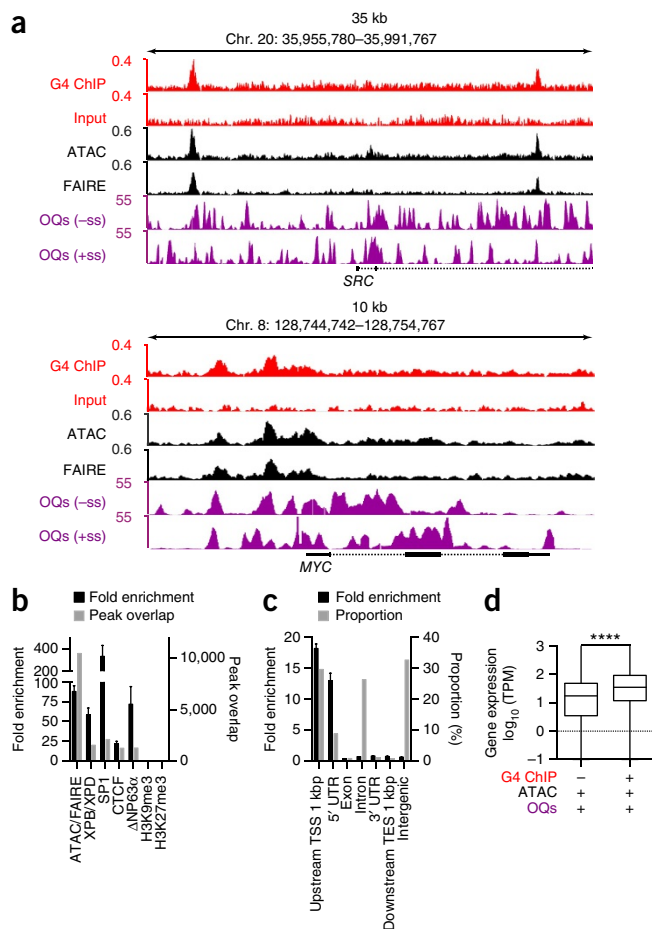


Figure 1 G4 sites are prevalent in regulatory chromatin regions. **(a)** Example genome browser screenshots for *SRC* and *MYC*. Tracks are shown for HaCaT G4 ChIP-seq (top, red) and control input (bottom, red); regulatory chromatin sites from ATAC-seq and FAIRE-seq (black); and G4-seq peaks (pyridostatin-derived observed quadruplex sequences (OQs); percent mismatches in reads aligned) on the reverse (–ss) and forward (+ss) strand¹⁶ (purple). Genomic coordinates indicate track range. **(b)** Enrichment and peak overlap of HaCaT G4 ChIP-seq peaks with different ENCODE protein ChIP-seq data sets⁴³. Error bars, s.d. ($n = 3$). **(c)** Enrichment of HaCaT G4 ChIP-seq peaks relative to what is expected by randomly shuffling G4 ChIP-seq peaks in OQs and the absolute proportion of G4 ChIP-seq peaks across different genomic features. Error bars, s.d. ($n = 3$). TES, transcription end site; TSS, transcription start site. **(d)** The distribution of mRNA levels (displayed in transcripts per million (TPM), log₁₀ scale) is shown for genes (4,345) with promoters that feature an ATAC-seq peak and an OQ signature (median = 1.2; interquartile range (IQR) = 0.4–1.7) in comparison to genes (4,522) that exhibit a promoter G4 ChIP-seq peak, ATAC-seq peak and OQ feature (median = 1.6; IQR = 1.1–2.0) in HaCaT cells. **** $P < 0.0001$, unpaired two-tailed t -test.

microscopy colocalization for BG4 and another G4 antibody (1H6) (ref. 14) with euchromatin and heterochromatin markers in HaCaT cells. We found that both G4 antibodies significantly ($P < 0.05$) colocalized with transcriptionally active euchromatin (signaled by the presence of trimethylation of histone H3 at lysine 4 (H3K4me3) and/or RNA polymerase II (RNA Pol II)) but did not show significant ($P = 0.2$) colocalization with heterochromatin (signaled by trimethylation of histone H3 at lysine 9, H3K9me3) (Supplementary Fig. 4). Collectively, these results demonstrate that G4 formation in human cells is predominantly restricted to regulatory, nucleosome-depleted regions in euchromatin.

Overall, HaCaT G4 ChIP-seq peaks represented about 1% of all sequences detected by G4-seq¹⁶ (hereafter referred to as observed quadruplex sequences, or OQs) (Supplementary Fig. 5). However, of all the OQs that mapped to nucleosome-depleted regions in HaCaT cells, only a subset (26%) overlapped with G4 ChIP-seq peaks (Supplementary Fig. 5). Importantly, the remaining 74% of OQs that were not detected by BG4 in nucleosome-depleted regions showed a representation of G4 motifs (sequence and structure) comparable to that of the subset detected by G4 ChIP (26%) (Supplementary Fig. 5). This suggests that additional genomic features or events beside nucleosome depletion are likely to be important for stable formation of G4 structures in chromatin. While the G4 ChIP-seq peaks were considerably enriched—i.e., found more often than expected by random chance—in promoter and 5' UTRs, they were mostly located outside these regions (Fig. 1c). This observed enrichment raises the possibility that the transcriptional state of a gene might affect G4 formation in nucleosome-depleted regions. Indeed, we observed that genes displaying a G4 ChIP-seq peak in their promoter (4,522) had, on average, significantly ($P < 0.0001$) higher transcriptional levels than genes lacking a promoter G4 ChIP-seq peak (4,345) yet that were still found to be nucleosome depleted and contain an OQ (Fig. 1d). As regulatory proteins shape open chromatin, we anticipated that published ChIP-seq peaks for XPB, XPD¹ or SP1 (ref. 29), which unwind or bind G4 structures *in vitro*, or for ΔNP63α, a master regulator of keratinocyte transcription³⁰, would be enriched in the G4 ChIP-seq data. Indeed, a notable enrichment of all XPB and XPD (60-fold), SP1 (330-fold) and ΔNP63α (72-fold) high-confidence peaks was observed in the HaCaT G4 ChIP-seq data, suggesting that these proteins might directly interact with endogenous G4 structures (Fig. 1b). In contrast, and as anticipated, the ChIP-seq peaks for H3K9me3 and trimethylation of histone H3 at lysine 27 (H3K27me3) showed no correlation with the G4 ChIP-seq peaks (Fig. 1b). Taken together, these results suggest that the chromatin context restricts G4 formation to predominantly regulatory, nucleosome-depleted regions associated with genes showing elevated transcription. This suggestion is consistent with transcriptional upregulation of predicted G4-forming genes observed in human cells deficient for G4 helicases (e.g., WRN and BLM)³¹ and with the observation of binding sites for the G4 helicases XPB and XPD and the yeast PIF1 homolog Pfh1 in transcriptionally active chromatin^{1,32}. Figure 1a shows example profiles for G4 ChIP-seq, ATAC-seq and FAIRE-seq aligned with peak profiles that mark OQs for the *SRC* and *MYC* oncogenes, which have previously been suggested to be regulated by G4 structures^{2,10}. Here we directly confirm the presence of G4 structures in the nuclease hypersensitivity element of *MYC* and in the upstream element and gene body of *SRC* (Fig. 1a). Recently, we reported enrichment of OQs at cancer-related genes and/or at somatic copy number alterations (SCNAs)¹⁶; we now extend these findings to demonstrate G4 (i.e., G4 ChIP) enrichment in cancer-related genes, such as *MYC*, *TP53*, *JUN*, *HOXA9*, *FOXA1* and *RAC1* (Supplementary Fig. 6 and Supplementary Table 1), and in SCNAs (Supplementary Fig. 7 and Supplementary Table 2). Among all cancer-related SCNA amplifications and oncogenes, *MYC* showed the highest G4 ChIP density, supporting a previously suggested role³³ for G4 structures in cancer progression (Supplementary Figs. 6 and 7).

We reasoned that, if G4 structures are coupled to nucleosome depletion and transcriptional status, then changing the chromatin landscape would cause a concomitant shift in the G4 profile. We induced chromatin relaxation in HaCaT cells, using the histone deacetylase (HDAC) inhibitor entinostat³⁴ to stabilize transcriptionally active chromatin through promoter-specific acetylation of histone H3K27 (ref. 35). HaCaT cells treated with 2 μM entinostat

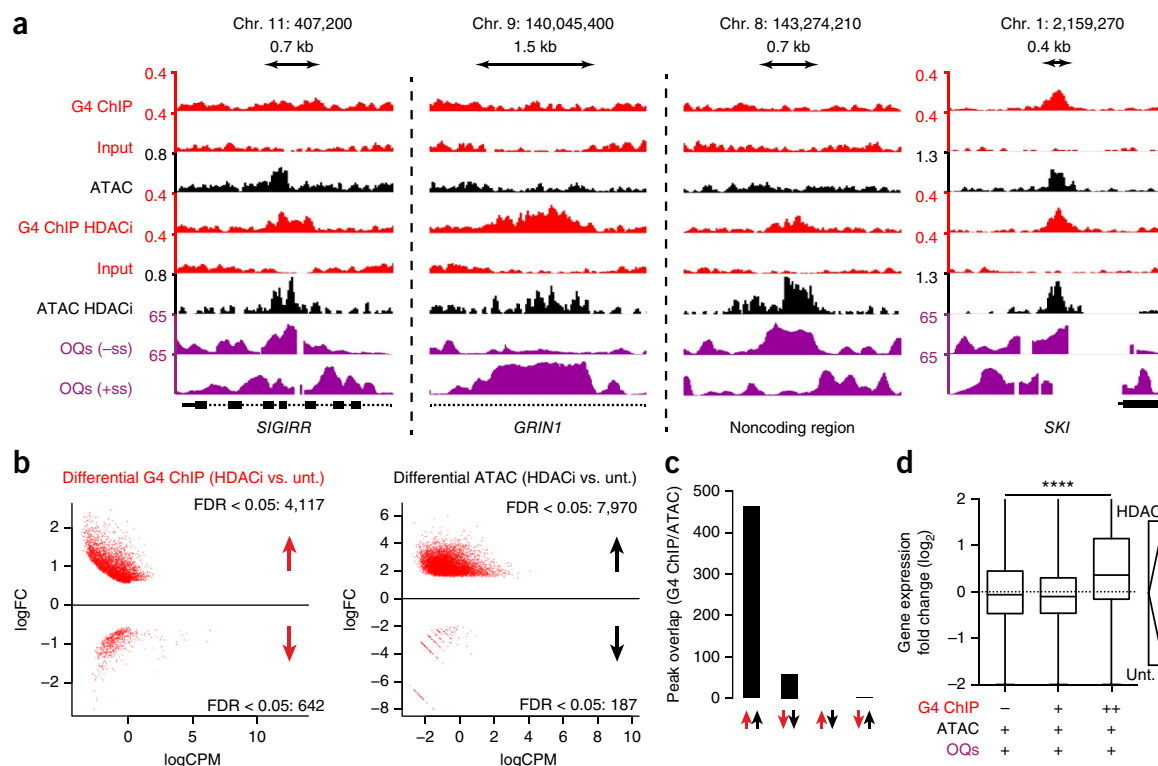


Figure 2 Chromatin relaxation increases G4 prevalence in regulatory chromatin regions. **(a)** Genomic view of *SIGIRR*, *GRIN1* and a noncoding region showing overlap between new G4 sites and regulatory chromatin in entinostat (HDACi)-treated cells. The *SKI* promoter (right) exemplifies a gene that was not significantly changed in G4 ChIP-seq or ATAC-seq peak size. Tracks are (from top to bottom) G4 ChIP-seq and input for untreated HaCaT cells (red); ATAC-seq for untreated HaCaT cells (black); G4 ChIP-seq and input for entinostat-treated HaCaT cells (red); ATAC-seq for entinostat-treated HaCaT cells (black); and OQs (PDS derived; purple) for the reverse (–ss) and forward (+ss) strand¹⁶. Genomic coordinates indicate peak summits in G4 ChIP HDACi tracks. **(b)** DBA showing differences in G4 ChIP-seq and ATAC-seq peaks with FDR < 0.05 between entinostat-treated and untreated (unt.) HaCaT cells. Red dots represent peaks where G4 ChIP or nucleosome depletion (ATAC-seq) is significantly changed in entinostat-treated as compared to untreated HaCaT cells. Red and black arrows indicate increase or decrease of G4 formation and changes in nucleosome depletion (ATAC-seq), respectively. FC, fold change; CPM, read count per million. **(c)** Overlap between G4 ChIP-seq peaks with increased or decreased signal under entinostat treatment and open chromatin regions, derived from **b**. **(d)** Differential expression for genes with (i) promoters that do not contain high-confidence G4 ChIP-seq peaks but do have at least one OQ and ATAC-seq peak unaltered in size across conditions (1,734 genes), (ii) promoters as in (i) but that contain at least one high-confidence G4 ChIP-seq peak shared across conditions and do not contain G4 ChIP-seq peaks significantly increased in size under entinostat treatment (3,627 genes), or (iii) promoters as in (i) but that contain G4 ChIP-seq peaks significantly increased in size under entinostat treatment (373 genes). **** $P < 0.0001$, ordinary one-way ANOVA.

for 48 h³⁵ were analyzed by G4 ChIP-seq, ATAC-seq and RNA-seq (Fig. 2 and Supplementary Fig. 8). Genome-wide changes in G4 ChIP-seq peaks, open chromatin and mRNA levels were quantified bioinformatically by differential binding analysis (DBA)³⁶ (Fig. 2a,b and Supplementary Fig. 8). HDAC inhibition resulted in the appearance of 4,117 new, or more intense, G4 ChIP-seq sites and 7,970 open chromatin regions (Fig. 2a,b). The 4,117 emergent G4 ChIP-seq sites are located in new or pre-existing nucleosome-depleted regions, such as the *SIGIRR* and *GRIN1* genes and a noncoding region (Fig. 2a,c). Importantly, however, none of these sites were found in nucleosome-depleted regions in closed chromatin after entinostat treatment (Fig. 2a,c). Next, we explored the relationship between G4 sites identified in nucleosome-depleted promoters and the transcriptional status of associated genes by comparing RNA-seq data to G4 ChIP-seq peaks for untreated and entinostat-treated cells. Consistent with the outcome in untreated cells (Fig. 1d), we observed that genes with OQs in their promoters that overlapped with both a G4 ChIP-seq peak and an ATAC-seq peak had, on average, significantly higher transcriptional output ($P < 0.0001$) than genes whose promoters lacked a G4 ChIP-seq peak in entinostat-treated HaCaT cells (Supplementary Fig. 8). Importantly, for genes with promoters showing a new or larger

G4 ChIP-seq peak but no significant change (\log_2 (fold change) = -0.6 to 0.6 , FDR < 0.05) in open chromatin state after entinostat treatment, transcriptional output also increased in comparison to the corresponding genes in the untreated state ($P < 0.0001$) (Fig. 2d). This indicates that there is a positive and dynamic relationship between G4 structure and transcriptional activity, independent of the degree of chromatin accessibility (Fig. 2d), and further suggests that G4 structures epigenetically mark the genome, whereby four-stranded structure formation, rather than the underlying sequence *per se*, is directly linked to elevated transcription.

We next investigated how an altered cellular state results in chromatin, G4 and transcriptional changes by comparing normal human epidermal keratinocytes (NHEKs) and their spontaneously immortalized counterparts, HaCaT cells. Using G4 ChIP-seq, ATAC-seq and FAIRE-seq (Fig. 3 and Supplementary Fig. 9), we found that, while NHEKs exhibited 85,668 more nucleosome-depleted regions than HaCaT cells (Supplementary Fig. 9), there were actually fewer G4 ChIP-seq peaks in NHEK as compared to HaCaT cells (1,496 versus 10,560) (Fig. 3b). The G4 ChIP-seq peaks that were identified in HaCaT cells (8,478; 80%), but were absent from NHEKs, were located entirely within nucleosome-depleted regions common to both cell

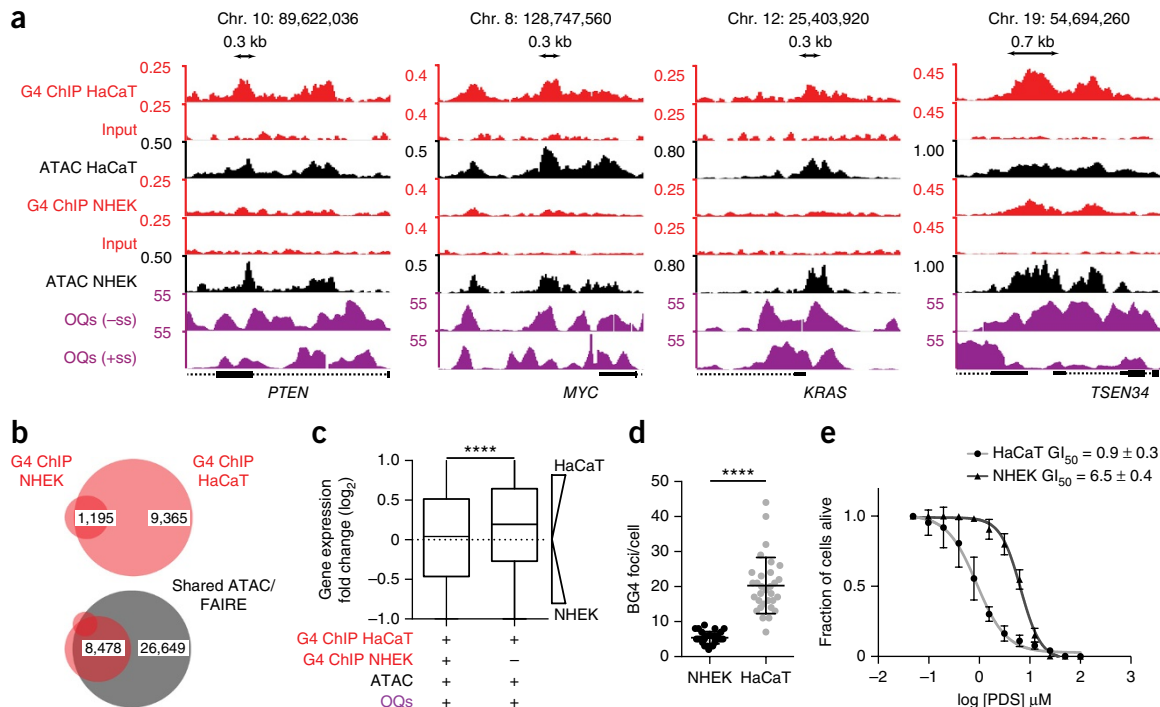


Figure 3 G4 prevalence is significantly increased in immortalized cells as compared to normal human epidermal keratinocytes. **(a)** Genomic view of *PTEN*, *MYC*, *KRAS* and *TSEN34* in HaCaT and NHEK cells. Tracks (from top to bottom) are G4 ChIP-seq and input for HaCaT cells (red), ATAC-seq for HaCaT cells (black), G4 ChIP-seq and input for NHEKs (red), ATAC-seq for NHEKs (black), and OQs (PDS derived; purple) on the reverse (–ss) and forward (+ss) strand¹⁶. Genomic coordinates indicate peak summits in G4 ChIP HaCaT tracks. **(b)** Overlap of G4 ChIP-seq peaks between HaCaT cells and NHEKs (top), and overlap between G4 ChIP-seq peaks and shared open chromatin regions (ATAC/FAIRE) (bottom). While both NHEK and HaCaT cells share G4 and active chromatin regions, NHEKs have ~7-fold fewer G4 sites. **(c)** mRNA levels for genes featuring (i) a promoter with a G4 ChIP and ATAC-seq peak and an OQ signature (503 genes; median = 0.04, IQR = –0.5 to 0.5) common to NHEK and HaCaT cells and (ii) a promoter with a unique G4 ChIP/ATAC/OQ peak signature in HaCaT cells as compared to NHEKs (3,617 genes; median = 0.19, IQR = –0.27 to 0.64). **** $P < 0.0001$, unpaired two-tailed t -test. **(d)** Quantification of immunolocalization (Supplementary Fig. 9) showing that the average number of BG4 foci per cell increases in HaCaT cells as compared to NHEKs. 109 and 333 HaCaT nuclei and 186 and 326 NHEK nuclei were analyzed in two replicate experiments ($n = 2$). Error bars, s.d.: **** $P < 0.0001$, unpaired two-tailed t -test. **(e)** Growth inhibition response curves for treatment with PDS ($n = 6$; error bars, s.d.). The concentration to obtain 50% growth inhibition (GI_{50}) is indicated for HaCaT cells (light gray curve) and NHEKs (dark gray curve).

lines (Fig. 3a). This suggests that additional mechanism(s), beyond having a nucleosome-depleted environment, control G4 formation in NHEK relative to HaCaT cells. Indeed, genes comprising promoter G4 ChIP-seq peaks with OQs that were present exclusively in HaCaT cells showed an overall increase in transcription as compared to genes containing promoter G4 ChIP-seq peaks common to both cellular states (Fig. 3c and Supplementary Fig. 9).

We identified G4 sites in many cancer genes, such as *MYC*, *PTEN* and *KRAS*, in immortalized HaCaT cells but not in NHEKs (Fig. 3a), suggesting a link between increased proliferative capacity/immortalization²³ and G4 structures. We determined whether the increased presence of G4 sites in HaCaT cells versus NHEKs was also reflected in the G4 foci observed in nuclei, as visualized by BG4 immunofluorescence microscopy¹³. Consistent with our G4 ChIP-seq data, we found that HaCaT cells showed ~4-fold (20 ± 8 versus 5 ± 2 ; errors are s.d.) more G4 foci than NHEKs (Fig. 3d and Supplementary Fig. 9). We have previously shown that the G4-selective small molecule pyridostatin (PDS) binds G4 structures, inducing a transcription- and replication-dependent DNA damage response and thereby inhibiting cell growth^{10,37}. We therefore investigated the PDS sensitivity of HaCaT cells as compared to NHEKs. We found that HaCaT cells were ~7 times more sensitive to growth inhibition by PDS than NHEKs, as determined by 50% growth inhibition (GI_{50}) values (0.9 ± 0.3 versus 6.5 ± 0.4 μ M), consistent with the observed

increase in the number of G4 structures in HaCaT cells (Fig. 3e). This increased abundance of G4 sites mirrors our previous observation that G4 structures can be more prevalent in cancer than in matched normal tissue³⁸ and suggests a potential rationale for selective cancer intervention by G4 targeting.

Here we provide the first high-resolution genome-wide map of G4 structures in human chromatin. The number of G4 structures (~10,000 G4 ChIP-seq peaks) is substantially lower than was predicted by computation²⁶ or observed by G4-seq¹⁶, and likely reflects the generally suppressive role of heterochromatin for G4 formation in human cells (for example, nucleosome density), which may be different in other species³⁹. More G4 ChIP-seq peaks are observed than BG4 immunofluorescence foci¹³ owing to the higher sensitivity and resolution of the ChIP-seq method. We conclude that G4 sites show hallmarks of dynamic epigenetic features in chromatin primarily found in regulatory, nucleosome-depleted regions and correlate with genes showing elevated transcription (Fig. 4). While small-molecule G4 stabilization can promote a DNA damage response^{10,40} and can cause transcriptional repression², our study suggests that endogenous G4 structures in promoters are linked to elevated transcriptional activity. We have also discovered that the G4 landscape is dynamically altered in a way that depends on chromatin relaxation or cell status, and that G4 motifs are particularly enriched in cancer-related genes and regions predisposed to amplification in cancer. Our study further illustrates

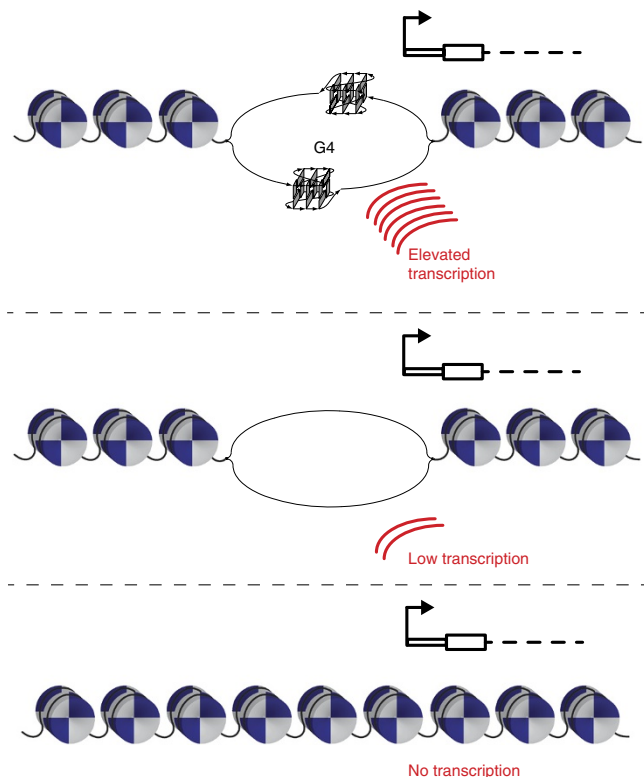


Figure 4 G4 DNA formation in chromatin. G4 DNA formation is highly dependent on chromatin structure and is frequently found in regulatory, nucleosome-depleted regions in proximity to the transcription start sites of genes that undergo elevated transcription.

the potential of the G4 structural motif as a molecular target for disease, diagnosis and intervention.

URLs. A sample sheet describing the detailed experimental design is available at <https://github.com/sblab-bioinformatics/dna-secondary-struct-chrom-lands>. Details of data analysis have been deposited at <https://github.com/sblab-bioinformatics/dna-secondary-struct-chrom-lands>. Lists of oncogenes and tumor-suppressor genes were obtained, respectively, from the COSMIC database (<http://cancer.sanger.ac.uk/census>)⁴¹ and tumor-suppressor gene database (<http://bioinfo.mc.vanderbilt.edu/TSGene/>)⁴². Peak correlations were performed using the BEDTools package (<http://bedtools.readthedocs.org>).

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Accession codes. The data reported in this paper are available at the NCBI GEO repository under accession number [GSE76688](#).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The authors would like to thank the staff at the Genomic and Light Microscopy and Biorepository core facilities at Cancer Research UK Cambridge Institute. We are grateful to the European Molecular Biology Organization for funding R.H.-H. with the EMBO Long-Term Fellowship. We acknowledge support from University of Cambridge and Cancer Research UK program. The Balasubramanian and Narita laboratories are supported by core funding from Cancer Research UK (C14303/A17197). S.B. is a Senior Investigator of the Wellcome Trust (grant no. 099232/z/12/z).

AUTHOR CONTRIBUTIONS

R.H.-H. developed the G4 ChIP-seq method with help from S.V.L. R.H.-H. carried out all experiments except for the immunofluorescence microscopy and growth-inhibition experimental work and analysis. K.Z., A.P. and M.D.A. carried out immunofluorescence microscopy experiments. M.D.A. performed growth inhibition experiments. R.H.-H., D.B. and G.M. designed, implemented and performed the bioinformatic analysis. R.H.-H., D.B., S.V.L., D.T. and S.B. designed epigenome experiments. J.P. performed analysis and quantification for colocalization immunofluorescence microscopy experiments. H.K. provided the antibodies to H3K9me3 (clone CMA304), H3K9me3 (clone CMA318) and RNA Pol II C-terminal domain (clone CMA601). All authors analyzed and interpreted the results. R.H.-H. wrote the manuscript with support and contributions from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

- 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).
- 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).
- Bochman, M.L., Paeschke, K. & Zakian, V.A. DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **13**, 770–780 (2012).
- Ribeyre, C. *et al.* The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo. *PLoS Genet.* **5**, e1000475 (2009).
- Cheung, I., Schertzer, M., Rose, A. & Lansdorp, P.M. Disruption of dog-1 in *Caenorhabditis elegans* triggers deletions upstream of guanine-rich DNA. *Nat. Genet.* **31**, 405–409 (2002).
- Haeusler, A.R. *et al.* C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507**, 195–200 (2014).
- Maizels, N. G4-associated human diseases. *EMBO Rep.* **16**, 910–922 (2015).
- Wolfe, A.L. *et al.* RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* **513**, 65–70 (2014).
- Rodriguez, R. & Miller, K.M. Unravelling the genomic targets of small molecules using high-throughput sequencing. *Nat. Rev. Genet.* **15**, 783–796 (2014).
- Rodriguez, R. *et al.* Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.* **8**, 301–310 (2012).
- Thurman, R.E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
- Stergachis, A.B. *et al.* Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. *Cell* **154**, 888–903 (2013).
- Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **5**, 182–186 (2013).
- Henderson, A. *et al.* Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.* **42**, 860–869 (2014).
- 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).
- Chambers, V.S. *et al.* High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* **33**, 877–881 (2015).
- Besnard, E. *et al.* Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat. Struct. Mol. Biol.* **19**, 837–844 (2012).
- Fouk, M.S., Urban, J.M., Casella, C. & Gerbi, S.A. Characterizing and controlling intrinsic biases of lambda exonuclease in nascent strand sequencing reveals phasing between nucleosomes and G-quadruplex motifs around a subset of human replication origins. *Genome Res.* **25**, 725–735 (2015).
- Hershman, S.G. *et al.* Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **36**, 144–156 (2008).
- 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).
- 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).
- Kanoh, Y. *et al.* Rif1 binds to G quadruplexes and suppresses replication over long distances. *Nat. Struct. Mol. Biol.* **22**, 889–897 (2015).
- Boukamp, P. *et al.* Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. *J. Cell Biol.* **106**, 761–771 (1988).
- Hogan, G.J., Lee, C.K. & Lieb, J.D. Cell cycle-specified fluctuation of nucleosome occupancy at gene promoters. *PLoS Genet.* **2**, e158 (2006).
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).

26. Huppert, J.L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33**, 2908–2916 (2005).
27. Guédin, A., Gros, J., Alberti, P. & Mergny, J.L. How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.* **38**, 7858–7868 (2010).
28. Mukundan, V.T. & Phan, A.T. Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. *J. Am. Chem. Soc.* **135**, 5017–5028 (2013).
29. 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).
30. McDade, S.S. *et al.* Genome-wide analysis of p63 binding sites identifies AP-2 factors as co-regulators of epidermal differentiation. *Nucleic Acids Res.* **40**, 7190–7206 (2012).
31. Johnson, J.E., Cao, K., Ryvkin, P., Wang, L.S. & Johnson, F.B. Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential. *Nucleic Acids Res.* **38**, 1114–1122 (2010).
32. Sabouri, N., Capra, J.A. & Zakian, V.A. The essential *Schizosaccharomyces pombe* Pfh1 DNA helicase promotes fork movement past G-quadruplex motifs to prevent DNA damage. *BMC Biol.* **12**, 101 (2014).
33. 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).
34. Saito, A. *et al.* A synthetic inhibitor of histone deacetylase, MS-27-275, with marked in vivo antitumor activity against human tumors. *Proc. Natl. Acad. Sci. USA* **96**, 4592–4597 (1999).
35. Tomazou, E.M. *et al.* Epigenome mapping reveals distinct modes of gene regulation and widespread enhancer reprogramming by the oncogenic fusion protein EWS-FLI1. *Cell Rep.* **10**, 1082–1095 (2015).
36. Ross-Innes, C.S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 (2012).
37. 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).
38. 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).
39. Hoffmann, R.F. *et al.* Guanine quadruplex structures localize to heterochromatin. *Nucleic Acids Res.* **44**, 152–163 (2016).
40. Gomez, D. *et al.* Telomestatin-induced telomere uncapping is modulated by POT1 through G-overhang extension in HT1080 human tumor cells. *J. Biol. Chem.* **281**, 38721–38729 (2006).
41. Forbes, S.A. *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **43**, D805–D811 (2015).
42. Zhao, M., Kim, P., Mitra, R., Zhao, J. & Zhao, Z. TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. *Nucleic Acids Res.* **44**, D1023–D1031 (2016).
43. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

ONLINE METHODS

Cell culture. HaCaT cells were kindly provided by F. Watt (King's College London Centre for Stem Cells and Regenerative Medicine, London, UK) and cultured in Dulbecco's Modified Eagle Medium (Thermo Fisher, DMEM, cat. no. 41965-039) supplemented with 10% FBS (Thermo Fisher). U2OS cells were cultured in DMEM (Thermo Fisher, cat. no. 41966-029) supplemented with 10% FBS. Normal human epidermal keratinocytes, pooled from multiple donors, were purchased from Thermo Fisher and cultured in EpiLife medium supplemented with human keratinocyte growth supplement (HKGS) (Thermo Fisher). Cell line genotypes were certified by the supplier or by STR profiling. Cells lines were confirmed mycoplasma-free by RNA capture ELISA.

G-quadruplex ChIP-seq protocol and library preparation. 20 million cells were fixed in DMEM (cat. no. 41965-039) containing 1% (v/v) formaldehyde and 10% (v/v) FBS for 15 min at room temperature (RT) followed by quenching with 0.13 M glycine at RT for 10 min. Cells were washed in ice-cold PBS and chromatin was isolated and prepared as described by the manufacturer (Chromatrap). Chromatin was sonicated using a Bioruptor Plus (Diagenode). 250 μ L of lysed nuclei suspension was sonicated for 25 cycles (30 s on/60 s off) in 1.5 mL TPX sonication tubes (Diagenode) to an average DNA size of 100–500 bp. For chromatin immunoprecipitation (ChIP), 2.5 μ L chromatin was blocked at 16 °C for 20 min in 44.5 μ L intracellular salt solution (25 mM HEPES pH 7.5, 10.5 mM NaCl, 110 mM KCl, 130 nM CaCl_2 , 1 mM MgCl_2), containing 1% (w/v) bovine serum albumin (BSA) (Sigma, cat. no. B4287), referred as blocking buffer. Recombinant BG4 and control phage display antibody were expressed as described¹⁴. 200 ng of recombinant BG4 or control phage display antibody were then added and incubated for 1 h at 16 °C. Meanwhile, 90 μ L beads (Sigma-Aldrich Anti-FLAG M2 Magnetic Beads, cat. no. M8823) were washed 3 \times with 900 μ L blocking buffer and stored on ice in 900 μ L blocking buffer (bead blocking solution). Next, 50 μ L of bead blocking solution was added to the ChIP reaction (final volume 100 μ L) and incubated for 1 h at 16 °C with rotation. Beads were magnetically captured and the supernatant was discarded, and the beads were then washed three times in 200 μ L wash (4 °C) buffer (10 mM Tris pH 7.4, 100 mM KCl, 0.1% (v/v) Tween 20) with manual agitation. Beads were resuspended in 200 μ L wash buffer and incubated on a rotating platform at 37 °C for 10 min. The warm wash procedure was repeated, and beads were subjected to a final wash (4 °C) before removal of the supernatant and resuspension in 75 μ L of elution buffer, containing 1 \times TE buffer with 20 μ g proteinase K (Thermo Fisher). Elution was performed at 37 °C for 1 h and at 65 °C for additional 2 h. Beads were magnetically captured, and eluted DNA was purified from supernatant using a MinElute kit (Qiagen). For each technical replicate, eluted DNA from four ChIP reactions were combined and the pool subjected to Nextera library preparation as described by the manufacturer (Illumina, cat. no. FC-121-1030). Three independent technical replicates were pulled-down and sequenced for each of the two biological replicates.

G-quadruplex ChIP-qPCR. Purified and sonicated DNA (as above) was used to quantify G4 enrichment via qPCR, using Fast SYBR PCR mix (Applied Biosystems, UK), with a Bio-Rad CFX96 quantitative PCR machine. Cycling conditions were 95 °C for 20 s followed by 40 cycles of 3 s at 95 °C and 30 s at 60 °C. We employed primer pairs that target G4 ChIP positive and negative regions (**Supplementary Table 3**). Relative enrichments were derived with respect to their inputs and normalized to a G4-free enhancer region associated with the *ESR1* gene (**Supplementary Table 3**). RNase A has been shown to digest G4 RNA^{13,44,45} and G4 DNA/RNA hybrids⁴⁵. For DNase or RNase A treatments, 2 U of TURBO DNase (2 U/ μ L; Thermo Fisher Scientific) or 1 μ g of RNase A (1 mg/mL, Ambion Thermo Fisher Scientific) were added to the ChIP blocking buffer chromatin mixture and chromatin digestion, instead of chromatin blocking, was performed at 37 °C for 20 min.

Epigenome mapping. FAIRE-seq was performed essentially as described, with minor modifications⁴⁶. Cell lysis was performed using a Precellys 24 homogenizer, and phenol-chloroform extraction was performed using Heavy Phase Lock Gels (5 PRIME). Transposase-accessible chromatin using sequencing (ATAC-seq) was performed essentially as described, with minor modifications²⁵. Tagmented DNA samples were amplified (11 PCR cycles)

using the Nextera Index Kit (Illumina, cat. no. FC-121-1011) and open chromatin fractions were size selected (190–300 bp) via 2% (w/v) agarose E-Gel SizeSelect gels (Thermo Fisher). Total RNA for RNA-seq experiments was extracted using the RNeasy kit (Qiagen, cat. no. 74104), following the manufacturer's instructions. RNA-seq libraries were generated using the Illumina TruSeq RNA HT (stranded mRNA) kit (cat. no. RS-122-2103). Overall, 18 G4 ChIP-seq, 18 ATAC-seq, 4 FAIRE-seq and 12 RNA-seq libraries were made (**Supplementary Fig. 1**). All epigenome libraries were sequenced in single-end or paired-end using 75-bp reads and the NextSeq 500 platform.

Data analysis. Mapping, peak calling and peak processing. Raw FASTQ reads were trimmed to remove adaptor contamination and aligned to the human reference genome version hg19 using cutadapt⁴⁷ and bwa⁴⁸, respectively. Reads with a mapping quality below ten or marked as positional duplicates were excluded from further analysis. FAIRE-seq, ATAC-seq and G4-ChIP peaks were mapped using MACS2.0 (ref. 49). RNA-seq reads were aligned to the human reference genome with TopHat2 (ref. 50). The Galaxy Cistrome platform⁵¹ was employed to process MACS2.0-called peak intervals and to generate high-confidence peak overlaps between biological replicates and area-proportional Venn diagrams. Find individual motif occurrences (FIMO)⁵² analysis was used to quantify the density of the multiple Em for motif elicitation (MEME)⁵³ motifs in the high-confidence G4 ChIP-seq data (HaCaT).

Differential gene expression. Gene expression levels were quantified as transcripts per million (TPM). Genes differentially expressed between (i) HaCaT cells treated with entinostat and untreated HaCaT and (ii) NHEK vs. HaCaT cells were identified using the Bioconductor package edgeR (FDR < 0.05; fold change > 1.5) (ref. 54).

G4 transcriptional analysis. Promoter transcription start site (TSS) coordinates, 1 kb (+/–) from TSS, were generated for 22,483 genes using hg19. To identify absolute gene expression values (TPM) for promoters with and without a G4 ChIP-seq signature, we divided promoters into two categories: (i) promoters that overlap with at least one high-confidence ATAC-seq peak, at least one OQs (potassium- and PDS-induced) and that overlap with at least one high-confidence G4 ChIP-seq peak, and (ii) promoter regions as in (i) that do not overlap with the high-confidence G4 ChIP-seq data set of untreated HaCaT cells (command intersect, subtract BEDTools package).

Differential gene expression levels for promoter-associated genes were prepared into three subsets as follows: Promoters that (i) do not overlap with G4 ChIP-seq peaks, but have at least one OQs and ATAC-seq peak unaltered in size (\log_2 fold change = –0.6 to 0.6, FDR < 0.05 differential binding analysis³⁶) between untreated and entinostat-treated HaCaT cells, (ii) are as in (i) but do not overlap with G4 ChIP-seq peaks significantly increased in size for entinostat-treated vs. untreated HaCaT cells and overlap with at least one high-confidence G4 ChIP-seq peak shared between untreated and entinostat-treated HaCaT cells, or (iii) are as in (i) but overlap with G4 ChIP-seq peaks significantly increased in size (\log_2 fold change = > 0.6, FDR < 0.05 differential binding analysis³⁶) for entinostat-treated vs. untreated HaCaT cells (command intersect, subtract BEDTools package).

Epigenome enrichment analysis. ChIP-seq data from the ENCODE project was retrieved from the NCBI's GEO repository as follows: XPB and XPD (GSE44849); SP1 (ENCSR991ELG; ENCSR000BJX; ENCSR000BHK); H3K9me3 (ENCSR000EYF; ENCSR000ARN; ENCSR000APE); H3K27me3 (ENCSR000EWB; ENCSR000DWU; ENCSR000DUE); CTCF (ENCSR000DWX; ENCSR000EGM; ENCSR000DUG); DeltaNp63 (GSE32061). Overlap between the high-confidence HaCaT G4 ChIP-seq peak file and the ENCODE ChIP-seq data sets was tested using the Galaxy Cistrome platform⁵¹. Common high-confidence ENCODE ChIP-seq peak files across the three different cell lines were calculated, if applicable (see GEO accession numbers). ENCODE ChIP-seq peak files were randomly shuffled ($N = 6$) across the genome and potential overlaps were tested with the G4 ChIP-seq peak file (command shuffleBed of the BEDTools package). Enrichments between G4 ChIP-seq and ENCODE ChIP-seq data sets were calculated from the ratio of the direct overlaps with the randomly shuffled overlaps.

G4 ChIP-seq peak annotation and enrichment analysis. PAVIS⁵⁵ was used to annotate the HaCaT G4 ChIP-seq peaks. The G4 ChIP-seq file was randomly shuffled across a file containing all OQs (potassium and PDS induced) (command shuffleBed of the BEDTools package) three independent times and their

genomic annotations were retrieved via PAVIS. To calculate enrichments of annotated features, ratios of the annotated G4 ChIP-seq proportions were performed with the proportions of the randomly shuffled features ($N = 3$).

G4 motif and enrichment analysis. Sequences within peak regions were analyzed by regular expression matching and assigned to one of the following classes: Loop size 1–3, 4–5 and 6–7; sequences with at least one loop of the respective length; long loop: sequences with a G4 with any loop of length >7 (up to 12 for any loop and 21 for the middle loop); simple bulge: sequences with a G4 with a bulge of 1–7 bases in one G-run or multiple 1-base bulges; 2-tetrads/Complex bulge: sequences with a G4s with two G-bases per G-run or several bulges of 1–5 bases; and other: other G4 types that do not fall into the former categories. Sequences in each peak were assigned to these classes in a hierarchical fashion; for instance, if a sequence matched both the Loop 1–3 and Long loop categories, it was assigned to the former category, in this case Loop 1–3. The fold-enrichment analysis was calculated by comparing actual counts for each class within peak regions to counts of the same peak regions after random reshuffling throughout the genome using the BEDTools shuffle command. Results are shown as the average of 10 randomizations, and the fold enrichment bar plot displays the ratio of real counts divided by average random counts. For fold enrichment analysis of G4 motifs compared to similar motifs with lower G4 forming potential, we compared respectively: (i) G4 with 4 G-runs of at least 3 Gs each and loop size 1–7 (G_3L_{1-7}) compared to a similar motif with 3 G-runs of at least 3 Gs each and loop size 1–7; (ii) as in (i), but with loop length 1–12 (G_3L_{1-12}); (iii) as in (i), but with G-runs of 2 Gs each, i.e. 2-tetrads (G_2L_{1-7}); (iv) canonical G-quadruplex motifs as in (i) (G_3L_{1-7}) but with a single bulge of 1–7 bases ($GGH_{1-7}G$, with $H = A, T$ or C) in only one G-run compared to a similar motif with a single G ($H_{1-2}GH_{1-2}$) instead of the bulge motif; (v) same G-quadruplex as in (iii) (G_2L_{1-7}) compared to a similar motif with one G-run having a single G ($H_{1-2}GH_{1-2}$).

Cancer-related gene analysis. After remapping to hg19, 555 oncogenes and 1,211 tumor suppressors (Supplementary Table 1) were considered. The occurrence of ChIP-seq peaks in each gene was determined and compared to the occurrence within genes not annotated as cancer genes or as tumor suppressors. To account for different region size, the peak density was calculated as number of peaks per megabase (Mb). Peaks occurring within the gene body (i.e., exons + introns) and in proximal promoter regions (i.e., 3,000 bases upstream of the transcription start site) were counted separately. P values for statistical testing were calculated with the nonparametric Wilcoxon rank sum test (implemented as `wilcox.test()` function in the R software).

Somatic copy number alteration analysis. Somatic copy number alteration (SCNA) regions associated with cancer ($n = 108$), either amplifications ($n = 54$) or deletions ($n = 54$), (i.e., copy number gains or losses of any length and amplitude), were obtained from the Cancer Genome Atlas Pan-Cancer data set⁵⁶. ChIP-seq peak abundance in each region was compared to the occurrence of randomly reshuffled peak regions (5 different randomization) in a similar manner to the G4 motif enrichment analysis. To account for different region sizes, peak density was calculated as the number of peaks per megabase. For visual comparison, we also determined the average genome-wide peak density per megabase (average density = 3.91), calculated as total number of peaks divided by the effective genome size (2.7×10^9 for *hg19*), multiplied by 1,000,000 bp (Supplementary Table 2).

Immunofluorescence microscopy. BG4 immunostaining and fluorescence microscopy for the HaCaT and NHEK cells were performed as previously described¹³. All secondary or tertiary antibodies were obtained from Thermo Fisher unless otherwise stated. In brief, cells were fixed with 2% (v/v) formaldehyde in PBS, permeabilized with 0.1% (v/v) Triton X-100 in PBS, and blocked with 2% (w/v) Marvel milk (Premier Foods PLC) in PBS. After blocking, cells were incubated with BG4 followed by incubation with secondary Rabbit Anti-FLAG (cat. no. 2368, Cell Signaling Technology) and tertiary goat anti-rabbit conjugated with Alexa-594 (cat. no. A11037) at 37 °C in a humid chamber for 1 h each. Coverslips were mounted on Prolong Gold/DAPI Antifade (Thermo Fisher). Digital images were recorded using a DP70 camera (Olympus) on an Axioskop 2 Plus microscope (Zeiss) and analyzed with Fiji⁵⁷. 200–300 nuclei were counted per condition and s.d. calculated from

two biological replicates. Frequency distribution graphs were plotted using GraphPad Prism (GraphPad Software Inc.).

For BG4 or 1H6 colocalization studies with H3K4me3, H3K9me3 and RNA polymerase II, HaCaT cells were grown to 50% confluency on # 1.5 glass coverslips. Cells were fixed and permeabilized as described above, treated with 50 µg RNase A for 30 min at 37 °C and incubated in blocking buffer for 1 h at 37 °C (PBS/ 0.5% normal goat serum, 0.1% Tween 20). For BG4 studies, coverslips were incubated with BG4 (10 nM) for 1 h at 37 °C, washed three times for 5 min with PBST and incubated 37 °C for 1 h with rabbit Anti-FLAG 1:800 (cat. no. 2368, Cell Signaling Technology) and mouse anti-H3K9me3 1:200 (Clone CMA304) (ref. 58), or mouse anti-H3K9me3 1:200 (clone CMA318) (ref. 59), or mouse anti-RNA polymerase II C-terminal-domain (clone CMA601) 1:200 (ref. 60). For 1H6 studies, coverslips were incubated with mouse anti-quadruplex DNA (cat. no. Ab00389-1.1, Absolute Antibody) and rabbit anti-H3K9me3 1:200 (cat. no. 9751, Cell Signaling Technology) or rabbit anti-H3K4 me3 1:500 (cat.no. 07-523, Upstate) at 37 °C for 1 h. Subsequently, all coverslips were washed as described and incubated for 30 min at 37 °C with the following Alexa Fluor conjugated antibodies: goat anti-rabbit Alexa-488 (cat. no. A-11034) and goat anti-mouse Alexa-555 (cat. no. A-21424) for BG4 studies and goat anti-mouse Alexa-488 (cat. no. A-11001) and goat anti-rabbit Alexa-555 (cat. no. A-21429) for 1H6 studies. DAPI counterstain was included in the final antibody incubation. Following washing, all coverslips were mounted onto Superfrost Plus slides (cat. no. 4951PLUS, Thermo Fisher) with VECTASHIELD Antifade mounting medium (cat. No. H-1000, Vector Laboratories Ltd.). Three biological replicates were performed.

Confocal z-stack images (15 steps) were acquired using a Leica TCS SP8 microscope with a HC PL APO CS2 1.4NA 100× oil objective (Leica Microsystems), in bidirectional mode, at a scan speed of 400 Hz and sampling rate of $0.06 \mu\text{m} \times 0.06 \mu\text{m} \times 280 \mu\text{m}$. The 405 nm diode laser was used to excite the DAPI channel (at 405 nm) and the white-light pulsed laser (SuperK EXTREME, NKT Photonics) was used to excite secondary antibody fluorophores (at 555 nm and 488 nm). Fluorescence detection was performed in sequential acquisition mode with hybrid detectors (Leica HyD Photon Counter) at the collection wavelength ranges of 490–535 nm, 565–630 nm and 410–485 nm for Alexa Fluor 488, Alexa Fluor 555 and DAPI, respectively. The pinhole was set to one airy unit and laser power and gain settings were consistent between replicates. Five to six Z stacks were obtained per replica representing 60–160 cells each. Representative images were processed using ImageJ⁶¹/Fiji and Photoshop (Adobe Systems, Inc.) software and assembled using Adobe Illustrator CS4.

Colocalization analysis was performed using an automated workflow in MATLAB (R2015b, Mathworks Inc.) with a link to ImageJ through MIJ⁶². First, the nuclear region was isolated by blurring the DAPI channel with a 3D Gaussian (radius $0.2 \mu\text{m}$) and Otsu thresholding (stack histogram). Noise reduction of the signal of each colocalization channel was performed using the PureDenoise plugin (ImageJ)⁶³ and isolated by removing background with a rolling-ball approach (radius $2 \mu\text{m}$) and Otsu thresholding. The corrected first Mander's overlap coefficient ($M_{1\text{diff}}$) corresponds to the difference between the percentage of the G4 antibody (BG4 or 1H6) signal that colocalizes with euchromatin, heterochromatin or gene promoter signal; the expected value for randomly distributed signal was calculated on the isolated signal. Finally, a two-way t -test was performed for each condition across the mean values of $M_{1\text{diff}}$ across the three biological replicates⁶⁴, and corrected for multiple hypothesis testing using the Bonferroni correction ($n = 5$). Graphs were created using GraphPad Prism version 6.00.

Small-molecule treatment and growth-inhibition analysis. HaCaT cells were treated with 2 µM entinostat (E5477-5 mg, Cambridge Bioscience) for 48 h at 30% confluency in 150 mm tissue culture dishes before G4 ChIP fixation procedure or ATAC-seq preparation, or in 6-well plates to generate four technical replicates for RNA-seq preparation. The concentration to inhibit by 50% cell growth (GI_{50}) induced by PDS was quantified using the end point CellTiterGlo assay (Promega). Briefly, 4,000 HaCaT and NHEK cells were seeded in a 96-well white plates (Nunc) for 18 h before small-molecule treatment. Cells were then treated with PDS doses ranging from 50 to $0.012 \mu\text{M}$ for 72 h. Cells

were then incubated with Cell Titer Glow for 1 h at room temperature and luminescence was measured using a PHERAstar FS microplate reader (BMG LabTech). Cell survival curves were plotted and GI₅₀ values were calculated using GraphPad Prism (GraphPad Software Inc.). Error bars represent variability within six replicates.

44. 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).
45. 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).
46. Simon, J.M., Giresi, P.G., Davis, I.J. & Lieb, J.D. Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nat. Protoc.* **7**, 256–267 (2012).
47. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–11 (2011).
48. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <http://arxiv.org/abs/1303.3997> (2013).
49. Zhang, Y. *et al.* Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
50. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
51. Giardine, B. *et al.* Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* **15**, 1451–1455 (2005).
52. Grant, C.E., Bailey, T.L. & Noble, W.S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
53. Machanick, P. & Bailey, T.L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).
54. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
55. Huang, W., Loganantharaj, R., Schroeder, B., Fargo, D. & Li, L. PAVIS: a tool for Peak Annotation and Visualization. *Bioinformatics* **29**, 3097–3099 (2013).
56. Zack, T.I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).
57. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
58. Kimura, H., Hayashi-Takanaka, Y., Goto, Y., Takizawa, N. & Nozaki, N. The organization of histone H3 modifications as revealed by a panel of specific monoclonal antibodies. *Cell Struct. Funct.* **33**, 61–73 (2008).
59. Hayashi-Takanaka, Y. *et al.* Tracking epigenetic histone modifications in single cells using Fab-based live endogenous modification labeling. *Nucleic Acids Res.* **39**, 6475–6488 (2011).
60. Stasevich, T.J. *et al.* Regulation of RNA polymerase II activation by histone acetylation in single living cells. *Nature* **516**, 272–275 (2014).
61. Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
62. Sage, D., Prodanov, D., Tinevez, J.-Y. & Schindelin, J. in *ImageJ User & Developer Conference* 24–26 (2012).
63. Luisier, F., Vonesch, C., Blu, T. & Unser, M. Fast Interscale Wavelet Denoising of Poisson-Corrupted Images. *Signal Process.* **90**, 415–427 (2010).
64. McDonald, J.H. & Dunn, K.W. Statistical tests for measures of colocalization in biological microscopy. *J. Microsc.* **252**, 295–302 (2013).