Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing

Robert Hänsel-Hertsch[1,4], Jochen Spiegel[1,2,4], Giovanni Marsico[1], David Tannahill[1] & Shankar Balasubramanian[1–3]

[1]Cancer Research UK, Cambridge Institute, Li Ka Shing Centre, Cambridge, UK. [2]Department of Chemistry, University of Cambridge, Cambridge, UK. [3]School of Clinical Medicine, University of Cambridge, Cambridge, UK. [4]These authors contributed equally to this work. Correspondence should be addressed to S.B. (sb10031@cam.ac.uk).

Published online 22 February 2018; doi:10.1038/nprot.2017.150

G-rich DNA sequences can form four-stranded G-quadruplex (G4) secondary structures and are linked to fundamental biological processes such as transcription, replication and telomere maintenance. G4s are also implicated in promoting genome instability, cancer and other diseases. Here, we describe a detailed G4 ChIP-seq method that robustly enables the determination of G4 structure formation genome-wide in chromatin. This protocol adapts traditional ChIP-seq for the detection of DNA secondary structures through the use of a G4-structure-specific single-chain antibody with refinements in chromatin immunoprecipitation followed by high-throughput sequencing. This technology does not require expression of the G4 antibody *in situ*, enabling broad applicability to theoretically all chromatin sources. Beginning with chromatin isolation and antibody preparation, the entire protocol can be completed in < 1 week, including basic computational analysis.

INTRODUCTION

Chromatin structure is tightly linked to the (epi)genetic control of gene expression and replication. In particular, nucleosome-depleted regions (NDRs) present regulatory features within chromatin that enable the assembly of transcription and epigenetic machinery[1]. Although there are numerous chromatin regulators, ranging from transcription factors and chromatin-associated proteins to histone modifications and DNA methylation, alternative DNA secondary structures warrant serious consideration as features that regulate genome function[2].

The DNA G-quadruplex (G4) is a family of four-stranded secondary structures constructed from stacks of planar guanine-tetrads (G-tetrads) connected by intervening loop sequences. Computational approaches have been used to predict well over 300,000 putative quadruplex-forming sequences in the human genome[3]. G4s are enriched in key regulatory sites, including oncogene promoters and telomeres, as well as in gene bodies and 5′ UTRs[4]. G4s are stable under near-physiological conditions, and G4 targeting with small-molecule ligands can disrupt cellular processes[5–11]. A number of proteins recognize G4 structures with high affinity, including several helicases that show potent G4-unwinding activity[12]. A central goal in the field is to understand whether endogenous G4s act as incidental structures that must be resolved for normal cellular functions or whether G4s are elements that have a direct role in signaling and regulation[13–15].

Evidence has now accumulated in support of the endogenous existence of G4s (refs. 16–20), and there has been considerable debate regarding the nature of G4s in human cells and whether small-molecule ligands can target endogenous G4s (ref. 21). DNA G4s were first visualized in ciliates[16], then in human cells and tissues using immunofluorescence and immunohistochemistry with G4 structure–specific antibodies[17–19]. A recent *in vitro* map of G4s in single-stranded, purified human genomic DNA revealed >700,000 G4 structures (called observed quadruplex sequences, OQs)[22], which was considerably greater than early predictions because of inclusion of structural variations such as fewer tetrads, longer loops and/or the presence of non-G-tetrad bases (bulges)[23] that widen the scope of G4 structures beyond earlier definitions. To probe G4 formation in chromatin, we developed G4 chromatin immunoprecipitation-sequencing (G4 ChIP-seq). While taking inspiration from ChIP-seq for transcription factor and chromatin-binding proteins, G4 ChIP-seq has been optimized for detection of the DNA G4 structure by using the BG4 scFv G4-antibody[17] to immunoprecipitate G4 structures from fixed chromatin. Recently, we reported the application of G4 ChIP-seq to create a genome-wide map of endogenous G4 structures in the chromatin of cultured primary and spontaneously immortalized human keratinocytes[24]. These experiments reveal that G4s primarily mark regulatory NDRs, and are strongly associated with highly transcribed genes. Therefore, G4s may represent epigenetic features that are intimately involved in chromatin function.

Comparison with other methods

An early attempt to map G4s used a DNA-immunoprecipitation approach with an antibody called hf2, using purified, fragmented human DNA, and uncovered a small set (~300) of stable G4-forming regions[25]. More recently, G4-seq revealed >700,000 G4s in single-stranded, purified human genomic DNA[22]. This map was generated using G4-promoting conditions to stimulate DNA polymerase stalling at G4s during Illumina/Solexa sequencing. Although they provide an experimental reference map of the potential of G4 formation, these experiments do not account for the endogenous effects of chromatin on G4 formation.

A different approach to detecting G4s in cells exploited the mapping of double-strand break sites induced by G4 stabilization with small-molecule ligands[11]. However, this approach was low in resolution (>kilobases) because of the spread of the γH2AX ChIP-seq signal. In related work, the resolution of mapping G4-related strand breaks has been improved by ChIP-seq of RAD51, which is more focally activated[26]. Although these overall approaches are suitable for exploring G4 targeting by small molecules,

they depend on chemical intervention and DNA damage induction, and are therefore not suitable for evaluating normal aspects of G4 function.

G4 formation has been inferred from ChIP-seq experiments that use antibodies against proteins with *in vitro* G4-binding activity. Enrichment of sequence motifs predicted to assume a G4 structure was uncovered for the human X-linked SWI/SNF protein, ATRX[27], and XPB/XPD helicases[20], yeast Rif1 (ref. 28) and the Pif1 (ref. 29) helicase. This is very supportive of G4 formation *in vivo*, but this may detect only specific G4s for the individual protein being tested, and as these proteins may bind other nucleic acid sequences or structures and proteins, the overall picture may be more complex. Protein ChIP-seq may be able to identify additional G4 sites that are unavailable in G4 ChIP-seq because of the BG4 antibody epitope being masked by the bound protein, and could therefore be a complementary approach.

Recently, an antibody, D1, displaying restricted structural specificity for parallel G4s, was expressed in cells as a GFP-fusion protein for ChIP-seq[30]. However, expression of this antibody-GFP fusion protein may lead to competition with endogenous proteins that bind the G4 motif, disturbing endogenous chromatin function and hence artificially perturbing the endogenous G4 landscape.

In comparison with conventional ChIP protocols for proteins, there are several key features that are required for the precipitation of G4 DNA secondary structures: (i) the application of a G4 structure-specific single-chain (scFV) antibody, (ii) an optimized immunoprecipitation protocol to simultaneously suppress nonspecific chromatin binding and G4 RNA or G4 DNA–RNA hybrid interactions, (iii) mild washing conditions that avoid harsh detergents but involve additional warm washes at 37 °C to remove residual nonspecific bound chromatin and (iv) the application of a modified Nextera library preparation protocol to convert a fragmented DNA population of heterogeneous size distribution to an unbiased homogeneously sized DNA population to allow the detection of heterochromatic regions and minimize a potential bias for NDRs.

Applications and outlook

ChIP-seq has revolutionized the understanding of genome function in basic biology and disease[31]. By far, the majority of such analyses rely on the availability of 'ChIP-seq grade' antibodies against the chromatin-binding protein/transcription factor in question, or against epitope tags engineered onto these proteins. For G4 structures, we have refined an approach that renders G4 structures amenable to ChIP-seq using one particular antibody, BG4. This approach suggests that there is the potential for other antibodies to be engineered to other structures (e.g., triplexes or cruciforms), which may also be amenable to a similar structure-based ChIP-seq methodology. Indeed, R-loops, consisting of RNA–DNA hybrids, have been mapped by ChIP-seq[32].

High-resolution (100–500 bp) G4 maps combined with maps of nucleosome positions, chromatin-associated proteins and transcription factors, as well as epigenetic marks, such as histone and cytosine modifications, offer the prospect of uncovering new regulatory insights. For example, our findings using a keratinocyte model system, show that G4 presence in NDRs is closely associated with increased transcriptional output of the adjacent gene. The G4 ChIP-seq method will therefore enable detailed investigations into the role of G4s in transcription and replication, both at the genome-wide level and at specific loci. Moreover, the 3D spatial relationship between G4 structures, nucleosome position and topologically associated domains may allow an understanding of how complex architectures organize different chromatin functionalities.

Overview of the procedure

Here, we present a ChIP-seq protocol for genome-wide identification of G4 DNA structure formation using the G4 structure-specific antibody BG4 (**Box 1**)[17]. To date, G4 ChIP-seq[24] has revealed (i) cell type and state formation for G4s in highly transcribed regions, (ii) *de novo* G4 formation after epigenetic reorganization and (iii) a greater number of G4s in immortalized versus primary cells. The protocol requires a relatively small number of cells (50,000–100,000) for a single G4 ChIP-seq experiment, is

Box 1 | BG4 antibody preparation and quality assessment

A high-quality antibody is essential for successful G4 ChIP-seq. We use the scFv antibody, BG4, which is prepared using the expression vector pSANG10-3F-BG4 (Addgene, plasmid no. 55756)[17].

BG4 expression ● TIMING 2 d

Grow BL21(DE3) *E. coli* carrying pSANG10-3F-BG4 at 37 °C in 2×YT medium (50 µg/ml kanamycin). Induce antibody expression at OD_{600} = 0.6 by addition of IPTG (0.5 mM) followed by incubation with shaking for 4 h at 37 °C. Pellet cells at 4,000*g* for 30 min at 4 °C and suspend the pellet in TES buffer (50 mM Tris-Cl, pH 8.0, 20% (wt/vol) sucrose) and incubate the bacterial slurry for 10 min on ice. Dilute the cells twofold in water and leave the suspension on ice for 10 min before centrifugation at 16,000*g* for 30 min at 4 °C. Collect the supernatant and load onto HIS-Pure cobalt spin columns (Sigma-Aldrich, cat. no. 89969), and then wash with 10 mM imidazole in PBS (pH 8.0). Elute the purified BG4 antibody in 250 mM imidazole in PBS (pH 8.0). Concentrate the purified BG4 solution using an Amicon Ultra-15 centrifugal filter unit (Millipore, cat. no. UFC9010) and dialyze against intracellular salt buffer. Snap-freeze the antibody (**Fig. 3a**) in small (30-µl) aliquots and store the aliquots at −20 °C; frozen BG4 aliquots are functional for several months.

BG4 quality assessment ● TIMING 2 d

Before G4 ChIP, validate BG4 affinity and specificity for G4s by standard ELISA, e.g., test the binding to a folded DNA G4 oligomer in comparison with a single-stranded DNA control (**Fig. 3b**)[17]. In addition, we suggest performing a preliminary pull-down assay using fluorescence-labeled G4 and control oligomers to qualitatively assess the enrichment of the G4 oligomers (**Fig. 3c**).

Briefly, perform Steps 16–26 using a 1:1 mix of labeled G4 oligomer (d(FITC-$G_3(T_2AG_3)_{11}T_2$), 10 fmol) and mutated oligomer (d(Texas Red-$GTG(T_2AGTG)_{11}T_2$), 10 fmol) and 1 µg of chromatin. Confirm immunoprecipitation of G4 structures via denaturing urea PAGE (**Fig. 3c**, lanes 5 and 6).

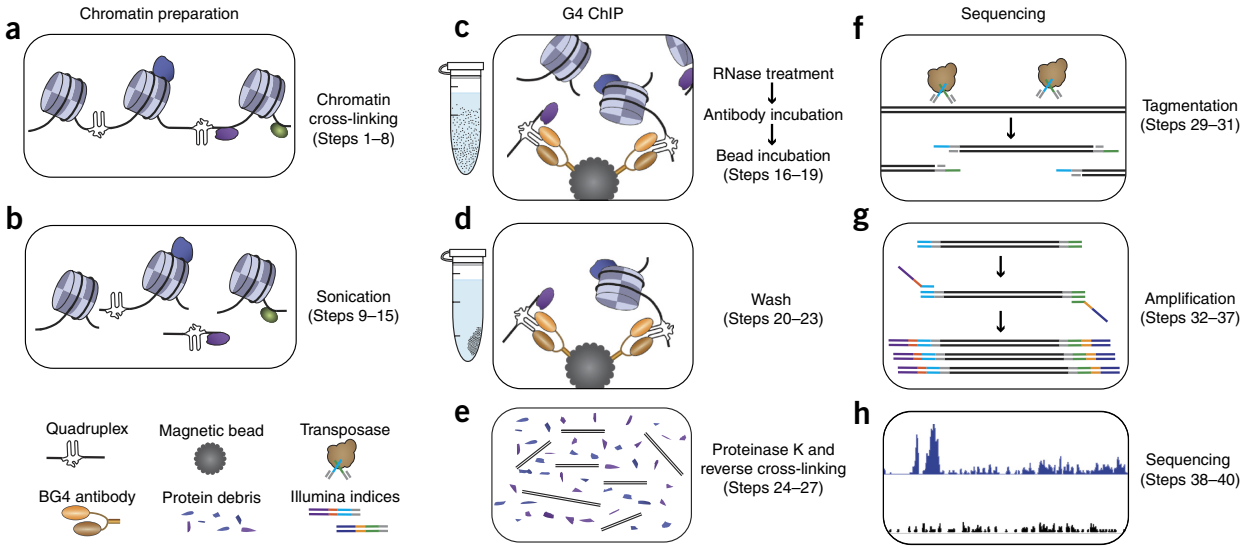

Figure 1 | General G4 ChIP-seq workflow. (**a**) G4 ChIP begins with cross-linking cells to stabilize DNA–protein interactions. (**b**) Nuclei are isolated, and chromatin is sheared by sonication. (**c**) Chromatin is treated with RNase A to remove RNA G4s and DNA–RNA G4 hybrids[50] and then is blocked with BSA to reduce nonspecific interactions. G4 structures are immunoprecipitated using the G4-specific antibody BG4 and immobilized on magnetic beads. (**d**) Beads are washed to reduce nonspecific interactions. (**e**) Immunoprecipitated DNA is eluted by reverse cross-linking and proteinase K treatment. (**f**) Isolated DNA is tagmented (simultaneous fragmentation and adaptor ligation via engineered transposomes). (**g**) PCR is used to amplify the library and add Illumina flow cell index adaptor sequences for multiplex sequencing. (**h**) Barcoded libraries are pooled together and subjected to next-generation sequencing.

amenable to parallel processing of many samples and generates robust genome-wide data across technical and biological replicates. The procedure has four main stages: (i) chromatin preparation (Steps 1–15); (ii) ChIP and quality assessment (Steps 16–28); (iii) library preparation (Steps 29–40) and (iv) sequencing and basic bioinformatics analysis (Steps 41–43). See **Figure 1** for a schematic illustration.

Chromatin preparation includes the following steps: cell preparation, cross-linking and washing, chromatin extraction via a two-step kit-based lysis protocol and sonication of chromatin to 100–500 bp. Chromatin handling, ChIP steps and their quality assessment are most critical and require chromatin and bead blocking, RNase treatment, G4 DNA capture, immobilization of BG4–G4 DNA complexes, multiple washes, elution of G4 DNA via proteinase K treatment and reverse cross-linking. Eluted and input DNA fractions are purified by solid-phase (MinElute/Qiagen) columns. G4 ChIP efficiency is validated using qPCR on enriched G4 DNA regions present in all cell types tested to date. Library preparation uses a modified Nextera tagmentation[33] (Illumina) procedure to generate library sizes (100–500 bp): eluted and purified ChIP fractions (1–10 ng) are tagmented using a lower concentration of Tn5 at 37 °C. Standard Nextera barcoding and Q5 High-Fidelity DNA Polymerase (NEB) are used to amplify tagmented DNA by PCR, and amplified libraries are cleaned using AMPure XP bead purification. Finally, average library size and quantity are calculated using a Bioanalyzer (Agilent) and Qubit fluorometer (Thermo Fisher Scientific).

Advantages and limitations of G4 ChIP-seq

G4 ChIP-seq detects chromatin regions that form G4 DNA structures in a cell population. As this is an enrichment-based method, it can detect G4 structures at a given locus present within a subset of cell states within a heterogeneous cell population. The approach also presents a snapshot in time and cannot resolve temporal dynamics of G4 formation. During the G4 ChIP-seq procedure, it is possible that BG4 binding to G4 structures may trap out structures that are in dynamic interconversion with other structural (or unstructured) states to favor the G4-folded structures over others. However, the observation that G4 ChIP-seq detects changes in G4 intensity and location in a way that appears to be dependent on the cell state or cell type argues against mapped G4 structures being a systematic artifact caused by antibody binding[24].

The current protocol does not resolve on which strand the G4 is formed, and the resolution is limited by the library fragment size (~100–500 bp), similar to most ChIP-seq protocols. Although our bioinformatics analysis of G4 ChIP sites did not reveal any BG4 bias toward certain G4 topologies, we cannot rule out G4 structures that are not recognized by the antibody during the procedure. The high affinity and broad selectivity of BG4 for a wide range of G4 structural types, including parallel, antiparallel and hybrid G4s (ref. 17), would appear to be advantageous for the protocol.

Some G4 DNA structures may be captured and enriched by BG4, but then lost during PCR amplification due to high GC content. Some G4s may be masked by endogenous proteins that cross-link to G4 DNA regions similar to an assembled histone octamer to B-DNA. As the majority of chromatin proteins cross-link with lower efficiency (≤1%) to DNA than histones, and in light of the high G4 ChIP input recovery rate (~10%), we suspect that BG4 mainly detects the protein-free G4 DNA state.

In any ChIP-seq experiment, heterochromatin, which is not generally associated with transcription factor binding, tends to be more resistant to shearing by sonication than euchromatin[34,35], and sonication may cause sample-specific biases resulting from different chromatin states. Furthermore, library preparation via fragment end-repair and ligation generates a population of long (> 500 bp) fragments enriched for heterochromatic regions that are less efficiently sequenced[34,35]. ChIP-seq data therefore

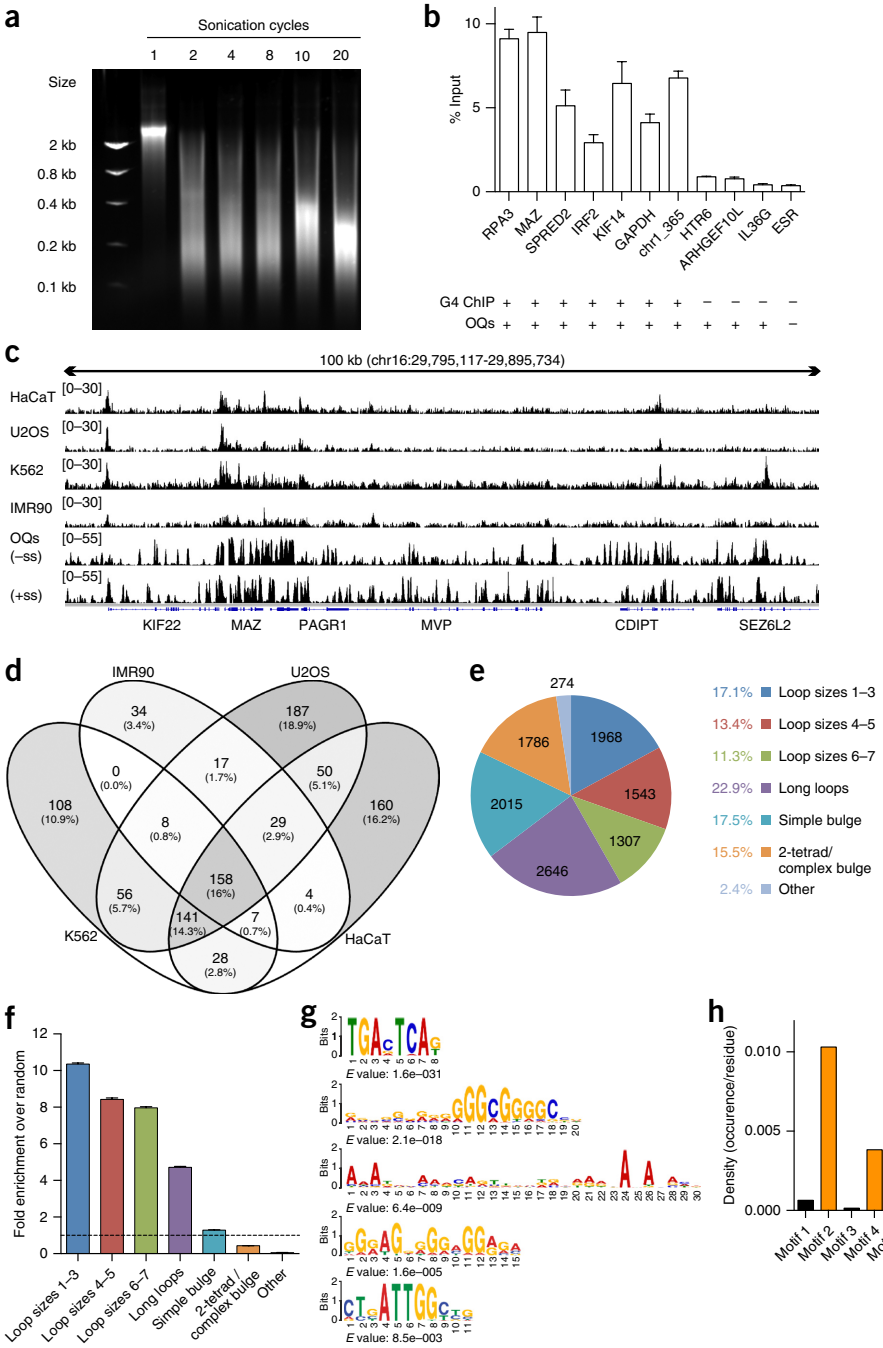

Figure 2 | Expected results from a typical G4 ChIP-seq experiment. (**a**) 1% (vol/vol) sheared chromatin from 2 × 10^7 HaCaT cells is displayed on a 2% (wt/vol) agarose EX E-gel. Chromatin preparations sonicated for 10 and 20 cycles are suitable for G4 ChIP-seq. (**b**) G4 ChIP-qPCR for selected G4 ChIP-seq peaks (**Table 1**) in HaCaT cells (error bars display s.d.; $N = 3$). Possible G4s (OQs)[22] are indicated if determined *in vitro* by G4-seq on the reverse (−ss) and forward (+ss) strands. (**c**) Example genome browser screenshot for a 100-kb region, showing common and differential G4 ChIP peaks in HaCaT, U2OS, K562 and IMR90 cell lines. Bottom tracks show possible G4s (OQs), as determined *in vitro* by G4-seq on the reverse (−ss) and forward (+ss) strands. G4 ChIP-seq data sets of chromosome 16 for all cell lines are available at the NCBI GEO repository under the accession number GSE99205. (**d**) Venn diagram showing the shared and unique G4 ChIP regions mapped in chromosome 16 of K562, U2OS, IMR90 and HaCaT cells. The larger the G4 ChIP fraction, the darker the background color. This diagram demonstrates that a substantial number of G4 ChIP sites are cell-type dependent and hence are dependent on the cellular state. (**e**) The total number of G4 ChIP-seq regions for HaCaT cells in each of the indicated structural classes of G4s are shown. Loop sizes 1–3, 4–5 and 6–7 indicate that at least one loop of this length is present in the G4; a long loop indicates a G4 with any loop of length >7 (up to 12 for any loop and 21 for the middle loop); a simple bulge indicates a G4 with a bulge of 1–7 bases in one G-run or multiple 1-base bulges; a 2-tetrads/complex bulge indicates G4s with two G-bases per G-run or several bulges of 1–5 bases; and other indicates other sequences that do not fall into the former categories. (**f**) Fold enrichment for each structural class in **e** compared with random (average of 10 randomizations). Higher enrichment values mean higher likelihood to be present among the G4 ChIP-seq peak sequences. Error bars indicate s.d. The dotted line indicates an enrichment of 1. (**g**) Motif discovery using MEME for a HaCaT G4 ChIP-seq peak data set reveals the presence of five possible recurring motifs (*E* value <0.1) within the entire peak file[39]. (**h**) Motif density (i.e., occurrence per residue) within the G4 ChIP-seq peak data set. Notably, only the G-rich motifs (highlighted in orange) show a considerable enrichment within the G4 ChIP-seq peak fraction. G4 ChIP-seq data sets for the HaCaT cell line are available at the NCBI GEO repository under the accession number GSE99205.

generally favors the preparation of shorter genomic fragments, which are also more enriched in NDRs. This raises the possibility that G4 ChIP-seq cannot detect G4s in heterochromatin. However, G4 ChIP-seq overcomes this through efficient fragmentation of chromatin into the 100- to 500-bp range (**Fig. 2**) and, critically, tagmentation converts large fragments into homogeneous DNA fragments for sequencing, thus recovering heterochromatin regions.

Future improvements

By reducing the volume during sonication, it should be possible to perform G4 ChIP-seq with fewer starting cells. Furthermore, chemical DNA footprinting or enzymatic processing, such as in ChIP-exo[36], of captured G4 DNA chromatin may improve the resolution of the G4 ChIP-seq method. Maps of G4 DNA structure formation in single cells may ultimately be enabled through microfluidic methods[37]. So far, G4 ChIP-seq has been used on a variety of different human cell lines (**Fig. 2**). We also expect that the protocol can be adapted to other sources of chromatin that include human tissues and other model organisms.

Experimental design

Cell number. A single G4 ChIP-seq experiment requires chromatin from 50,000 to 100,000 cells in a small volume (2.5 µl). To sonicate chromatin and optimize fragmentation, we find that it is optimal to prepare concentrated chromatin from 1 to 2×10^7 cells in a larger volume (500 µl of lysis buffer). If cell material is limited, it is feasible to use fewer cells in a smaller lysis buffer volume.

Cross-linking and cell lysis. We recommend fixing adherent and suspension cells with fresh methanol-free formaldehyde, using 1% (vol/vol) formaldehyde in growth medium containing 10% (vol/vol) fetal calf serum. As chromatin from different cells has different compactness and fragmentation properties, fixation duration can be optimized. Lysis conditions reported here have been successfully tested on a number of adherent and suspension cell lines.

Sonication. During sonication, chromatin samples heat up if not regulated by an automatically controlled cooling system. Overheating can lead to protein denaturation and can compromise chromatin quality. We use a Bioruptor Plus with 1.5-ml TPX microtubes (Diagenode) coupled to a permanently cooled water bath system. Per sonication cycle, chromatin is sonicated for 30 s with a 60-s recovery time between sonication cycles. The number of cycles needed to yield optimal fragmentation is variable and needs careful monitoring. We advise that 500 µl of chromatin be split into at least two tubes and that one tube be sonicated for 10 and the other for 20 cycles. We strongly recommend inspecting each chromatin preparation via gel electrophoresis using a 2% (wt/vol) agarose gel (see also **Fig. 2a**). Sonicated chromatin fractions are chosen on the basis of the least number of cycles sufficient to achieve an average chromatin fragmentation of 100–500 bp. We have not used other methods of DNA fragmentation, such as enzymatic digestion, as these potentially have more intrinsic bias, and it is not known how enzymatic cleavage will affect G4 DNA structure.

G4 ChIP. To generate enough immunoprecipitated DNA for library preparation, we previously combined three to four eluted and purified ChIP fractions from the original chromatin preparation. However, the product of a single G4 ChIP reaction yields enough input (1–10 ng) for Nextera library preparation. It is therefore sufficient to prepare a single G4 ChIP-seq library from a single G4 ChIP experiment and two additional technical replicates originating from the same chromatin batch, as well as input control; four samples in total. The used ratio of BG4 (0.2–0.5 µg) to chromatin (0.5–1 µg) has been optimized and robustly applied for several cell types (**Fig. 2c**), but may be refined further if DNA recovery rates are poor.

G4 ChIP enrichment quality assessment by qPCR. We recommend preparing at least three technical G4 ChIP replicates per biological experiment and two biological replicates. We use half of the eluted ChIP fraction to assess the quality of the G4 ChIP signal by G4 ChIP-qPCR; the other half is used for library preparation if enrichment quality is acceptable (**Box 2**). Using HaCaT, IMR90, K562 and U2OS cells, we have established several G4 DNA regions that show reproducible enrichment in the G4 ChIP-seq and G4 ChIP-qPCR data, for example, upstream of the *MAZ* gene (see **Fig. 2b,c** and **Table 1**)[24]. We advise setting up the G4 ChIP method via qPCR analysis using these G4 DNA regions (**Box 2**).

The use of input DNA for normalization. Analyzing sequencing results includes determining what is a true signal and what is random or nonspecific 'noise' generated through the G4 ChIP-seq process. We find that input DNA libraries display overrepresented (peak-like) DNA regions that must be accounted for when G4 ChIP-seq regions are bioinformatically identified. Therefore, the input sample is processed along with the ChIP sample, except for the immunoprecipitation steps (Steps 16–27). Hence, input libraries must be included in each experiment.

Library preparation and sequencing. Owing to its outstanding sensitivity in comparison with conventional library preparation methods, tagmentation has been recently used to generate low-input ChIP-seq libraries[38]. Applying an optimized tagmentation protocol, G4 ChIP-seq libraries can be prepared with ≥1 ng of immunoprecipitated DNA. Compared with the limited recovery of immunoprecipitated DNA (1–10 ng), chromatin input fractions contain much more DNA than immunoprecipitated DNA fractions. To minimize any potential PCR bias, we use 10 ng of input DNA for the tagmentation reaction, use standard Nextera barcoding and Q5 (NEB) to amplify ChIP and input libraries, and finally clean up libraries with AMpure XP beads. Owing to the

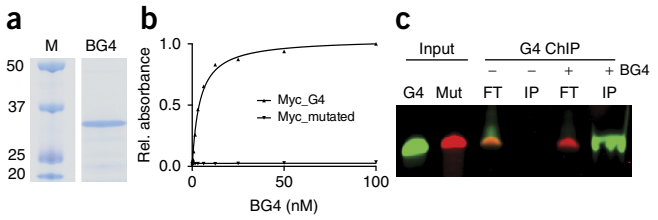

Figure 3 | BG4 quality assessment. (**a**) SDS PAGE showing molecular weight markers (M) and purified BG4 antibody (~31 kDa). (**b**) Representative ELISA testing binding of BG4 to MYC G4 and a mutated MYC sequence unable to fold into a G4 structure. (**c**) Urea PAGE analysis of a G4 ChIP control experiment using labeled G4 oligomer (G4, green) and a mutated control oligomer (Mut, red), showing flow-through (FT) and immunoprecipitated oligomers (IP).

Box 2 | qPCR analysis for assessing G4 DNA enrichment ● TIMING 2 h

For normalization purposes, the qPCR assay should always contain the input sample (fragmented DNA that was not immunoprecipitated), as well as several sets of primers directed against G4 ChIP-seq positive loci and at least one set of negative control primers. **Table 1** provides a selection of G4 ChIP-seq positive/negative primer sets that have been established in HaCaT cells and form robust G4 structures across three other cell lines (IMR90, K562 and U2OS). To obtain the fraction of recovery (% Input), normalize the results (Cq values) relative to the input sample (= $2^{(\text{input Cq}-\text{sample Cq})}$). To obtain an estimate for the G4 ChIP signal-to-noise ratio (fold enrichment), divide % Input of a G4-positive region by % Input of a G4-negative region. An acceptable G4 ChIP quality is achieved when the common G4 regions (e.g., *RPA3* and *MAZ*) yield DNA-input recoveries ≥5%, and their relative enrichment compared with G4-negative regions (e.g., *ESR1*, *TMCC1*) is ≥fivefold (see also **Fig. 2b**).

high complexity and low duplication rate (0–20%), G4 ChIP-seq libraries can be sequenced in single-end mode. To obtain high-quality G4 ChIP-seq maps, we advise a human genomic coverage of at least >0.7× (30 million clean reads; 75-bp read length in single-end mode) per library.

Bioinformatic analysis and validation of G4 ChIP-seq data. To identify G4 ChIP-seq regions, we use the input library as background files for MACS v2.0 peak calling, with default (q value = 0.05) or less-stringent settings (q value = 0.1). To reveal robust G4 ChIP regions common across all technical replicates, G4 ChIP regions from three technical replicates are intersected and filtered for overlaps. Generally, a robust peak is called when present in all three replicates. To determine the enrichment of the ChIP regions (.bed files) for G4 structures, we analyze different structural categories of computationally predicted G-quadruplex motifs within peak regions, and calculate the fold-enrichment for each class by comparing with random reshuffling across the genome (BEDTools shuffle command, ten randomizations; **Fig. 2e**,**f**), as previously described[24]. We further assess the overlap of the ChIP region with the *in vitro* biophysical map of experimentally determined G4s (OQs)[22]. To further validate the specificity of the ChIP for G-quadruplex motifs in an unbiased way, we interrogate the peak sequence data set for recurring motifs using the multiple EM for motif elicitation (MEME)[39] and find individual motif occurrences (FIMO)[40] packages (**Fig. 2g**,**h**). We find that, among the top motifs extracted by MEME, only G-rich sequences are found to be highly prevalent within the G4 ChIP-seq peaks.

MATERIALS

REAGENTS

- Cells (see Reagent Setup for details). HaCaT cells were kindly provided by F. Watt (King's College London Centre for Stem Cells and Regenerative Medicine). U2OS (ATCC, cat. no. HTB-96), K562 (ATCC, cat. no. CCL-243) and IMR90 (ATCC, cat. no. CCL-186) cells were purchased from ATCC ▲ **CRITICAL** Cell-line genotypes were certified by the supplier or by short tandem repeat (STR) profiling. Cell lines were confirmed to be mycoplasma-free by RNA capture ELISA. ! **CAUTION** The cell lines used in your research should be regularly checked to ensure that they are authentic and are not infected with mycoplasma.
- Formaldehyde, 16% (vol/vol), electron microscopy grade (Pierce; Thermo Fisher Scientific, cat. no. 28906) ! **CAUTION** Formaldehyde is toxic if inhaled, ingested or absorbed through the skin. Always wear a lab coat and gloves, and work in a chemical hood. All formaldehyde waste must be kept inside a chemical hood or in sealed containers, and collected for disposal by the proper authorities.
- Glycine, UltraPure (Thermo Fisher Scientific, cat. no. 15527013)
- BSA (Sigma-Aldrich, cat. no. B4287)
- ChIP-qPCR Kit (Chromatrap, cat. no. 500117) ▲ **CRITICAL** The chromatin extraction protocol of the commercial supplier has been used for convenience; however, an alternative chromatin extraction protocol has also been successfully tested[41]. Critically, G4 ChIP is sensitive to the presence of harsh detergents such as SDS or Triton X-100. We recommend not using >0.05% (wt/vol) SDS or >0.5% (vol/vol) Triton X-100 in the G4 ChIP reaction.
- Anti-FLAG M2 magnetic beads (Sigma-Aldrich, cat. no. M8823) ▲ **CRITICAL** Store at −20 °C, aliquot the beads and avoid repeated freeze–thaw cycles.
- Proteinase K (20 mg/ml; Ambion; Thermo Fisher Scientific, cat. no. AM2546). Store at −20 °C for up to several months
- RNase A (1 mg/ml; Ambion; Thermo Fisher Scientific, cat. no. AM2271). Store at −20 °C for up to several months
- Protease inhibitor cocktail solution (in ChIP-qPCR kit; Chromatrap, cat. no. 500117)
- MinElute Reaction Cleanup Kit (Qiagen, cat. no. 28206)
- SYBR Green PCR Master Mix (Applied BioSystems; Thermo Fisher Scientific, cat. no. 4312704). Store at 4 °C
- Recombinant BG4 (**Box 1**) ▲ **CRITICAL** The antibody can be stored at −20 °C for several months or for up to 2 d at 4 °C without substantial loss in activity. Avoid multiple freeze–thaw cycles.
- Expression vector pSANG10-3F-BG4 (Addgene, plasmid no. 55756)
- BL21(DE3) chemically competent cells (Sigma-Aldrich, cat. no. CMC0014)
- Kanamycin B sulfate (Sigma-Aldrich, cat. no. B5264)
- 2xYT medium (Sigma-Aldrich, cat. no. Y2377)
- IPTG (Sigma-Aldrich, cat. no. I6758)
- Sucrose (Sigma-Aldrich, cat. no. S9378)
- Phosphate-buffered saline (Sigma-Aldrich, cat. no. P4417)
- Imidazole (Sigma-Aldrich, cat. no. 792527)
- HIS-Pure cobalt spin columns (Sigma, cat. no. 89969)
- Nextera DNA Library Preparation Kit (Illumina, cat. no. FC-121-1030) ▲ **CRITICAL** Store at −20 °C for up to several months, aliquot reagents and avoid repeated thawing and freezing of buffers and enzyme.
- Nextera Index Kit (Illumina, cat. no. FC-121-1011). Store at −20 °C
- Agencourt AMPure XP beads (Beckman Coulter, cat. no. A63881). Store at 4 °C
- NEBNext High-Fidelity 2× PCR Master Mix (New England Biolabs, cat. no. M0541L). Store at −20 °C for up to several months
- Nuclease-free water (Ambion, cat. no. AM9932)
- Ethanol, pure (Sigma-Aldrich, cat. no. E7023-500 ml)
- Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, cat. no. Q32851)
- Agilent High Sensitivity DNA Kit (Agilent Technologies, cat. no. 5067-4626)
- Tween 20 (Sigma-Aldrich, cat. no. P2287)
- KCl (Sigma-Aldrich, cat. no. P9333)
- Liquid nitrogen
- G4 oligomers (Sigma-Aldrich)- Milli-Q-water (Millipore)

EQUIPMENT

- Water-cooled bath sonicator (Bioruptor Plus; Diagenode, cat. no. B02010002) **! CAUTION** Sonication causes damage to hearing. Wear ear protection. **! CAUTION** Pregnant women should not be exposed to sonication.
- TPX sonication microtubes, 1.5 ml (Diagenode, cat. no. C30010010-300)
- DNA/RNA LoBind microcentrifuge tubes, 1.5 ml (Eppendorf, Sigma-Aldrich, cat. no. Z666548-250EA)
- Rotating heat block (ThermoMixer C; Eppendorf, cat. no. 5382000.015)
- Refrigerated microcentrifuge (Eppendorf, model no. 5415R)
- CFX96 Touch real-time PCR detection system (BioRad, cat. no. 1855195)
- Magnetic separator (DynaMag-2; Thermo Fisher Scientific, cat. no. 12321D)
- Vortex mixer (Vortex Genie 2; Scientific Industries, cat. no SI-025)
- E1-ClipTip Electronic Single-Channel Pipette (Thermo Fisher Scientific, cat. no. 4670020)
- Qubit 3.0 Fluorometer (Thermo Fisher Scientific, cat. no. Q33216)
- Bioanalyzer 2100 (Agilent Technologies, cat. no. G2940CA)
- PCR tubes (Thermo Fisher Scientific, cat. no. 14230225)
- T100 thermal cycler (BioRad, cat. no. 1861096)
- Petri dishes (Sigma-Aldrich, cat. no. Z740356-80EA) or T-175 tissue culture flasks (Corning Life Sciences, cat. no. 431085)
- Electronic dispensing pipette (Thermo Fisher Scientific, cat. no. 4670020)
- Vacuum aspirator (Integra Biosciences, cat. no. 156400)
- Amicon Ultra-15 centrifugal filter unit (Millipore, cat. no. UFC9010)
- Cutadapt: https://github.com/marcelm/cutadapt
- BWA-MEM: https://github.com/lh3/bwa
- CleanSam tool: https://broadinstitute.github.io/picard/
- Picard tool: http://broadinstitute.github.io/picard/
- MACS v2.0: https://github.com/taoliu/MACS
- UCSC Genome Browser: https://genome.ucsc.edu/
- IGV: http://software.broadinstitute.org/software/igv/
- bamCoverage: https://github.com/fidelram/deepTools

REAGENT SETUP

Cross-linking solution Mix 1 ml of 16% (vol/vol) formaldehyde solution with 15 ml of growth medium containing 10% (vol/vol) serum. The solution should be prepared directly before use using a fresh ampule of formaldehyde.

2 M glycine Dissolve 37.6 g of glycine in Milli-Q water (may require gentle heating) and bring the volume to 250 ml with ddH_2O. The solution can be stored for at least a year at 20 °C.

Hypertonic buffer Chromatrap hypotonic buffer can be filter-sterilized using a 22-µm filter unit and then stored at 4 °C for several months. An aliquot should be supplemented with protease inhibitor cocktail solution directly before use.

Lysis buffer Chromatrap lysis buffer can be filter-sterilized using a 22-µm filter unit and then stored at 4 °C for several months. The buffer should be warmed to 40 °C for 30 min and then equilibrated to room temperature (20–25 °C). Add protease inhibitor cocktail solution directly before use.

TE buffer TE buffer is 10 mM Tris, pH 8.0, 1 mM EDTA in Milli-Q water. Filter-sterilize using a 22-µm filter unit. The buffer can be stored for at least a year at 20 °C.

Intracellular salt solution Intracellular salt solution is 25 mM HEPES, pH 7.5, 10.5 mM NaCl, 110 mM KCl and 1 mM $MgCl_2$ in Milli-Q water. Filter-sterilize using a 22-µm filter unit. The solution can be stored for at least a year at 20 °C.

Blocking buffer Blocking buffer is an intracellular salt solution containing 1% (wt/vol) BSA in Milli-Q water. Filter-sterilize using a 22-µm filter unit. The buffer can be stored in aliquots for up to a year at −20 °C.

Wash buffer Wash buffer is 100 mM KCl, 0.1% (vol/vol) Tween 20 and 10 mM Tris, pH 7.4, in Milli-Q water. Filter-sterilize using a 22-µm filter unit. The buffer can be stored for at least a year at 20 °C.

80% ethanol (vol/vol) Prepare fresh ethanol solution with nuclease-free water.

Cell culture preparation Cells should be cultured in medium and growth conditions that are appropriate for the given cell type.

! CAUTION *Mycoplasma* contamination tests should be carried out routinely, and cell identity should be confirmed by STR typing.

PROCEDURE

Growing and treatment of cells ● TIMING 12 h

1| Seed ~1 × 10^7 cells in 2 × 15 ml of growth medium in 2 × 15-cm Petri dishes or T-175 tissue culture flasks. Incubate the cells at 37 °C until the next day.

2| (Optional) Depending on the question being asked, at this point one can perturb the cells with the appropriate reagent; for example, we have previously used a histone deacetylase small-molecule inhibitor for 48 h to alter the epigenome and evaluate the dynamic changes to G4s in chromatin[24].

Formaldehyde cross-linking and harvesting of cells ● TIMING 1–2 h

3| For chromatin preparation of adherent cells, prepare 32 ml of a cross-linking solution by mixing 30 ml of growth medium with 2 ml of 16% (vol/vol) formaldehyde. Remove the growth medium from both dishes and add 16 ml of cross-linking solution to each of the dishes. Place the dishes on a rotary shaker inside a chemical hood and shake slowly at 30 r.p.m. Incubate the dishes for 10 min at room temperature. For suspension cell lines, transfer 1 × 10^7 cells to a 50-ml conical tube, pellet the cells by centrifugation at 300*g* for 5 min at 4 °C and replace the growth medium with 16 ml of fresh medium containing 1% (vol/vol) formaldehyde. Mix gently and incubate at room temperature for 10 min on a rotating platform.

▲ **CRITICAL STEP** Cross-linking time is critical to preventing chromatin overfixation and to minimizing extensive sonication. Do not cross-link for more than 10 min.

4| Stop the cross-linking reaction by adding glycine to a final concentration of ~0.12 M (1 ml of 2 M glycine per plate). Gently mix by rotating the reaction mixture for 10 min at room temperature.

5| Wash the cells twice with 10 ml of ice-cold PBS and add 10 ml of ice-cold PBS.

▲ **CRITICAL STEP** From this step onward, keep the samples and buffers ice-cold.

6| Scrape the cells and collect them in a 15-ml Falcon tube; one tube per plate.

TABLE 1 | Primer sets for G4-positive and G4-negative loci.

Primer name	Sequence (5′–3′)	G4 ChIP[a]	OQs
RPA3 forward	CGG AAG TTG ACA GAT ACA GGG	+	+
Reverse	GAT CGC AGA AAG GTA GTC TCA G		
MAZ forward	ACT CAG CGC AGG ATT GTA AAT A	+	+
Reverse	CCT CAT GCT TCG GCT TCC		
SPRED2 forward	AAC AGG AGG AGG AAG TAG GG	+	+
Reverse	TTT CGG TCG CAA GTA GGA AG		
IRF2 forward	TGA AAG CCC GTC AGT TGA ATA A	+	+
Reverse	GCT TTC GAT CTG GAC TGT TCTC		
KIF14 forward	CGG TAG CCG TCT CTG AAT G	+	+
Reverse	CTT TAG CAG AAC CCG AGG AG		
GAPDH forward	GCT ACT AGC GGT TTT ACG GGC G	+	+
Reverse	TGC GGC TGA CTG TCG AAC AGG		
Chr1_365 forward	TCT GGC GGC CGC TAT TG	–	+
Reverse	GCG GAC ACG GTT TGG ATA CTA		
HTR6 forward	GGC GAT TTG TCC AAT ATT TCC C	–	+
Reverse	CTG TGA CCT GCC CTT ATC C		
ARHGEF10L forward	TGC CAA GTT ACT CTC AGT TCT G	–	+
Reverse	AGC CAA ACC TCC AAG AAC AA		
IL36G forward	GCC CAC CTC TTT ACT TCC TTA	–	+
Reverse	AAC ACT CTT TCA GCT CCA TCC		
ESR1 forward	GAA ACA GCC CCA AAT CTC AA	–	–
Reverse	TTG TAG CCA GCA AGC AAA TG		
TMCC1 forward	GTG GTA CAC TGC CTA CAG TAT T	–	–
Reverse	GTA TAA CGC CTG GGC TAT GT		

[a]Based on the HaCaT G4 ChIP-seq landscape.

7| Pellet the cells by centrifugation at 300*g* for 5 min at 4 °C. Discard the supernatant.

8| Resuspend the cell pellet in 10 ml of ice-cold PBS and repeat Step 7.

■ **PAUSE POINT** At this point, cells can be flash-frozen and stored at 80 °C for up to 12 months.

Isolation and lysis of nuclei ● **TIMING 0.5–1 h**

9| If cells were frozen after Step 8, thaw the cells on ice before proceeding with lysis.

10| Resuspend the cell pellet in 0.8 ml of ice-cold Chromatrap hypotonic buffer per 2 × 10^7 cells. Incubate the mixture on ice for 10 min at 4 °C. Transfer the suspension to a 1.5-ml DNA LoBind tube. Centrifuge the mixture for 5 min at 5,000*g* at 4 °C and discard the supernatant.

11| Resuspend the nuclear pellet in 0.5 ml of Chromatrap lysis buffer and incubate on ice for 10 min.

Sonication and chromatin quality assessment ● TIMING variable

12| Split the nuclear suspension and transfer 0.25 ml to each of 2× 1.5-ml TPX sonication tubes. Using the Bioruptor Plus connected to a cooling system (4 °C), sonicate the samples under high-power mode for ten cycles (30 s on/60 s off). Keep the samples on ice.

13| Transfer sonicated chromatin from one of the two tubes to a fresh 1.5-ml TPX tube and repeat Step 12 for this tube only, keeping the remaining tube on ice.
▲ **CRITICAL STEP** TPX tubes are more fragile than other tubes, and not more than ten sonication cycles per tube are recommended.

14| To assess chromatin quality, spin the tubes at 21,000*g* for 10 min at 4 °C and then transfer the tubes to ice. Prepare two fresh 1.5-ml DNA LoBind tubes and add 70 µl of 1× TE buffer and 1 µl of RNase A to each tube. Add 5 µl of a sonicated sample to each prepared DNA LoBind tube, and incubate the tubes at 37 °C on a thermomixer for 20 min. Add 1 µl of proteinase K to each DNA LoBind tube, and incubate the samples in a thermomixer under rotation (1,400 r.p.m.) for 2 h at 65 °C. Clean up the DNA using the MinElute kit (Qiagen), following the manufacturer's instructions, and elute the samples in 20 µl of EB buffer (MinElute/Qiagen). Analyze chromatin fragment size distribution using 2% (wt/vol) agarose gel electrophoresis. If at least one of the chromatin samples shows a fragment distribution of 100–500 bp, proceed to Step 15; otherwise repeat Step 12 until appropriate size distribution has been achieved. Optionally, you can also quantify the global chromatin concentration via Qubit, according to the manufacturer's instructions. Typically, a chromatin concentration of 200–500 ng/µl is required for optimal G4 ChIP performance.
▲ **CRITICAL STEP** Sonication conditions may need to be optimized for different cell types and different equipment setups.
? TROUBLESHOOTING

15| Dispense chromatin in 25-µl aliquots in PCR tubes on ice and immediately snap-freeze in liquid nitrogen.
■ **PAUSE POINT** At this point, chromatin can be stored at 80 °C for up to 3 months.

G4 ChIP ● TIMING 2.5 h

16| For each G4 ChIP-seq chromatin batch/condition, prepare four 1.5-ml DNA LoBind tubes on ice, each containing 43.5 µl of blocking buffer, 1 µl of RNase A and 2.5 µl of chromatin. Incubate the four identical chromatin solutions in the thermomixer under rotation (1,400 r.p.m.) at 37 °C for 20 min.
▲ **CRITICAL STEP** This step enables sufficient enrichment of G4 DNA targets relative to background, such as non-G4 DNA or G4 RNA.

17| Add 200–500 ng of recombinant BG4 to three out of the four reactions, and incubate the three BG4-containing samples under rotation (1,400 r.p.m.) at 16 °C for 1 h; meanwhile, place the remaining tube on ice; it will serve later on as input control (Step 24).

18| Meanwhile, wash 65 µl of anti-FLAG beads three times with 650 µl of blocking buffer, resuspend the beads in 650 µl of blocking buffer and incubate side-by-side with the three BG4-containing reactions in the rotating thermomixer.

19| Add 50 µl of the beads in blocking solution from Step 18 to each of the three BG4-containing ChIP reactions (final volume = 100 µl) and incubate as in Step 17.

Washing of the beads ● TIMING 40 min

▲ **CRITICAL** From this step onward, to increase accuracy and reproducibility, we advise the use of an electronic dispensing pipette and a vacuum aspirator.

20| To wash the beads with 200 µl of ice-cold wash buffer, place the tubes with beads in a metal stand on a magnet to collect the beads; anti-FLAG beads are rapidly attracted to the wall of the tube adjacent to the magnet—to ensure an efficient bead wash, remove the chromatin reaction solution from the immobilized beads and discard. Add 200 µl of wash buffer to each tube, close the tubes, remove the plastic rack containing the tubes from the metal stand and invert the plastic rack several times to ensure proper bead wash. Finally, insert the plastic rack back into the metal stand and remove the supernatant. Repeat the addition and removal of 200 µl of wash buffer three times before proceeding to Step 21.
▲ **CRITICAL STEP** For all washing steps, work quickly and keep all wash buffers on ice. For best practices and reproducibility, perform alternating cycles of wash buffer addition and removal in rows for all tubes and work in a cold room (4 °C).

21| Add 200 µl of wash buffer to each tube, close and transfer the tubes to a hot (37 °C) rotating (1,400 r.p.m.) incubator for 10 min.

22| Immediately capture the beads, preferably in a cold-room, remove the supernatant and repeat Step 21.

23| Capture the beads at the tube wall using a magnet and make sure to remove any visible residual wash buffer before proceeding to Step 24.

DNA isolation ● TIMING 3 h

24| To elute G4 DNA from the beads, remove the tubes from the magnetic separator and add to each tube 75 µl of TE buffer and 1 µl of proteinase K. At the same time, fill the input tube saved from Step 17 with TE buffer to 75 µl and add 1 µl of proteinase K.

25| Transfer all four tubes to a thermomixer and incubate under rotation (1,400 r.p.m.) for 1 h at 37 °C and then for another 2 h at 65 °C.

■ **PAUSE POINT** At this point, a thermomixer can be programmed to hold the temperature at 16 °C overnight. Samples should be processed on the next day.

26| Capture the beads magnetically, retrieve the supernatant (eluted DNA) and purify the eluted DNA using a MinElute kit , following the manufacturer's instructions and eluting DNA in 25 µl of EB buffer.

27| To determine whether immunoprecipitated DNA is sufficient for Nextera library preparation, quantify 5 µl of the ChIP products and input via Qubit.

▲ **CRITICAL STEP** Most commonly, immunoprecipitated DNA yields are in the range of 2–10 ng, and all the material is used for a single Nextera tagmentation reaction. Importantly, an equivalent amount of DNA from the input fraction must be used in the tagmentation reaction (1–10 ng).

Assessment of G4 enrichment quality using qPCR ● TIMING 1.5 h

28| Assess G4 ChIP enrichment by G4 ChIP-qPCR, quantifying the relative enrichment of consistently enriched G4 DNA regions over background regions.

▲ **CRITICAL STEP** This analysis enables a first insight into the G4 ChIP-seq quality, as relative G4 ChIP-qPCR enrichments over fivefold generally lead to acceptable G4 ChIP-seq signal-to-noise ratios. We have observed consistent G4 structure formation in certain genomic regions, e.g., upstream of the *MAZ* gene, in the HaCaT, U2OS, K562 and IMR90 cell lines (**Fig. 2**, **Table 1** and as described in **Box 2**).

? TROUBLESHOOTING

Library preparation: tagmentation and adaptor ligation ● TIMING 1 h

▲ **CRITICAL** Library preparation is performed by tagmentation and makes use of the Nextera Library Preparation Kit (Illumina), which we have adapted as detailed below to accommodate the low DNA amount obtained in G4 ChIP-seq experiments.

29| To set up the tagmentation reaction, add 20 µl of TD buffer and 2.5 µl of TDE1 enzyme to 17.5 µl of DNA solution, mix on ice by repeated pipetting and centrifuge (280*g*, 1 min, 4 °C).

30| Incubate on a thermomixer for 20 min at 37 °C, rotating at 800 r.p.m.

31| Isolate the tagmented DNA using a MinElute kit, following the manufacturer's instructions, eluting the DNA in 21 µl of EB buffer.

32| Add 2.5 µl of each Nextera Index 1 (i7) and Index 2 (i5) adaptors, add 25 µl of NEBNext High-Fidelity 2× PCR master mix and transfer the mixture to a PCR tube; then mix by repeated pipetting and centrifuge (280*g*, 1 min, 4 °C).

33| Amplify the tagmented DNA using the following program.

Cycle number	Denaturation	Annealing	Extension	Termination
1	98 °C, 30 s			
2–12	98°C, 10 s	65 °C, 30 s	72 °C, 60 s	
13			72 °C, 5 min	4 °C, hold

Library preparation: library cleanup ● TIMING 30 min
▲ **CRITICAL** Steps 34–37 are performed at room temperature.

34| Centrifuge the condensate to the bottom of the PCR tubes, add 30 µl of AMPure XP beads to each tube and mix by repeated pipetting.

35| Incubate for 5 min, collect the beads on a magnetic separator and remove the supernatant and discard.
▲ **CRITICAL STEP** To avoid bead carryover or loss of beads, wait for at least 2 min for the beads to clear.

36| Wash the beads in each tube twice with 200 µl of 80% (vol/vol) ethanol at room temperature, discard the supernatant and keep the tubes open to dry on a magnetic separator for 10 min.
▲ **CRITICAL STEP** Keep the washing steps as short as possible. Complete removal of any liquid and droplets is essential. The beads must be completely dry before the next step.

37| Remove the PCR tubes from the magnetic separator and thoroughly resuspend the beads in 32.5 µl of Nextera RSB resuspension buffer per sample. Incubate the bead suspension for 2 min at room temperature, collect the beads on the magnetic separator and transfer 30 µl of the supernatant to a 1.5-ml LoBind reaction tube without disturbing the bead pellets.
■ **PAUSE POINT** The purified, tagmented DNA can be stored at −20 °C for at least a month.

Library preparation: quantification ● TIMING 1.5 h
38| Analyze 1 µl of each library on an Agilent 2100 Bioanalyzer system using an Agilent High Sensitivity DNA Kit, following the manufacturer's instructions to determine the size distribution of the library. Fragments of the sequencing library should have a size range of 150–600 bp.

39| Use 2 µl of each library to quantify the library concentration via Qubit, and convert the library concentration using the following table.

Library size from Bioanalyzer (bp)	Conversion factor from ng/µl to nM
250	1 ng/µl = 6 nM
500	1 ng/µl = 3 nM
1,000–1,500	1 ng/µl = 1.5 nM

? TROUBLESHOOTING

40| Normalize and pool the libraries (if multiplexing) according to the standard Nextera library preparation procedure in the manufacturer's instructions.

Sequencing and basic bioinformatics analysis ● TIMING 8 h
41| Sequence the libraries using a NextSeq or HiSeq Illumina platform.

42| *Sequencing quality and processing*. Evaluate sequencing quality by standard quality-control procedures[42]. Remove Illumina Nextera adaptor sequences using Cutadapt with default parameters[43]. Align trimmed sequences using BWA-MEM with default parameters[44]; discard sequences (in BAM file) with an alignment quality <10. Soft-clip local imperfect alignments using the CleanSam tool[45]. Finally, remove PCR duplicates by MarkDuplicates in the Picard tool using the default parameters. To obtain acceptable G4 ChIP-seq quality, we advise obtaining at least 30 million clean reads (in BAM format).

43| *Peak calling and genome browser track generation*. Run MACS v2.0 (ref. 46) with default parameters on the clean reads resulting from Step 42. Use the aligned reads from the immunoprecipitation as treatment and the reads from the input DNA as control. To create signal tracks to be uploaded to the UCSC[47] or IGV[48] genome browsers, generate bigwig files from the BAM files using the bamCoverage tool[49].

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 2**.

TABLE 2 | Troubleshooting table.

Step	Problem	Possible reason	Solution
14	Low chromatin quality	Inefficient lysis	Lyse the samples on ice for 10 min or longer, if required. Test alternative buffer setups[51], but avoid harsh detergents such as SDS or Triton X-100
		Sonication conditions used are not optimal for the specific cell type	Adjust sonication conditions. Reduce/increase the number of sonication steps and/or length of sonication cycles. Ensure that samples do not overheat during sonication. To resolve SDS in Chromatrap lysis buffer, make sure to warm and mix Chromatrap lysis buffer before use. DNA fragments after sonication should range from 100 to 500 bp
		Insufficient chromatin quantity	Increase the number of cells. In the case of insufficient sonication, increase the number of sonication steps
28	Low enrichment detected by qPCR after the G4 ChIP module	Low BG4 activity/signal-to-noise ratio	Assess BG4 activity by ELISA and ChIP with spiked fluorescently labeled G4 and single-stranded oligonucleotides (**Box 1**). Express a fresh batch of antibody, if necessary
		Cross-linking was not optimal (fixation time was too long, formaldehyde was not fresh, or concentration was too high)	Strictly adhere to cross-linking instructions. Use fresh formaldehyde. Do not use glycine solution older than 1 month
		Low chromatin quality and concentration	Optimize sonication conditions before G4 ChIP. Quantify the global chromatin concentration; it should be 200–500 ng/µl
		Problem with the reagents used for the ChIP module	G4 ChIP is sensitive to harsh detergents, such as SDS or Triton X-100. Prepare all reagents accurately
		Control regions are not optimal for the cell line	A particular G4 might not be prevalent in the individual cell line under investigation. Always use several positive and negative markers to assess G4 ChIP efficiency
	Excessive variation in qPCR data between replicates	Variable amounts of Anti-Flag magnetic beads between samples	Ensure that the beads are well suspended while pipetting. Avoid loss of beads during pipetting steps
39	Low library concentration	Amplification of tagmented DNA was not sufficient	Combine eluted fractions of multiple ChIPs to increase PCR template concentration
		The product is lost during cleanup steps	Ensure that ethanol traces evaporate during cleanup

● TIMING

Steps 1 and 2, growing and treatment of cells: 12 h
Steps 3–8, cross-linking and harvesting of cells: 1–2 h
Steps 9–11, isolation and lysis of nuclei: 0.5–1 h
Steps 12–15, sonication and chromatin quality assessment: variable
Steps 16–19, G4 ChIP: 2.5 h
Steps 20–23, washing of the beads: 40 min
Steps 24–27, DNA isolation: 3 h
Step 28, assessment of G4 enrichment quality using qPCR: 1.5 h
Steps 29–33, tagmentation and adaptor ligation: 1 h
Steps 34–37, library cleanup: 30 min
Steps 38–40, library quantification: 1.5 h

Steps 41–43, sequencing and basic bioinformatics analysis: 8 h
Box 1, BG4 expression: 2 d
Box 1, BG4 quality assessment: 2 d
Box 2, qPCR analysis for assessing G4 DNA enrichment: 2 h

ANTICIPATED RESULTS

Chromatin preparation

Correct chromatin shearing should result in fragments ranging from 100 to 500 bp. Depending on the cell line, cross-linking time and number of sonication cycles, fragments of other sizes may also be seen; however, the bulk should match this range, otherwise it will negatively affect subsequent steps (**Fig. 2a**).

G4 ChIP

Measurement of the DNA concentration at the end of the G4 ChIP module may give provisional evidence on the immunoprecipitation efficiency. DNA quantification (as measured by the Qubit HS DNA kit) depends on the number of cells used. For various cell lines, we typically obtain 1–10 ng of DNA from 50,000 to 100,000 cells. It is notable that DNA concentration is not indicative of the specificity of the DNA precipitated. Therefore, it is recommended to perform a qPCR assay (Step 31) to obtain preliminary data on the relative enrichment of G4-positive regions. Relative enrichments of at least fivefold generally lead to acceptable G4 ChIP-seq signal-to-noise ratios.

Library preparation

Quantification of the DNA amount after the library cleanup step is important for both evaluating the quality of ChIP-seq libraries and for pooling samples in correct ratios. The concentration is usually fivefold higher than the concentration after the ChIP module (10–50 ng).

Sequencing results

The number of detected G4 sites may vary depending on the specific cell line and treatment conditions. Examples of sequencing results from different cell lines are presented in **Figure 2**. We typically observe between 5,000 and 15,000 peaks and routinely obtain a reproducibility of >60% between biological replicates. We then determine the number (and position) of G4 ChIP-seq peaks that overlap with maps of potential G4s (OQs) and observe >80% overlap. Computational analysis identifies >90% of peaks containing one of the G4-predicted structural motifs (**Fig. 2e**,**f**).

ACKNOWLEDGMENTS We acknowledge support from the University of Cambridge and Cancer Research UK. The Balasubramanian laboratory is 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). J.S. is a Marie Curie Fellow of the European Union (747297-QAPs-H2020-MSCA-IF-2016).

AUTHOR CONTRIBUTIONS R.H.-H. developed the G4 ChIP-seq method. R.H.-H. and J.S. performed the experiments. R.H.-H., J.S. and G.M. performed bioinformatics analysis. All authors interpreted the data and wrote the manuscript.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

1. Thurman, R.E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
2. Bochman, M.L., Paeschke, K. & Zakian, V.A. DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **13**, 770–80 (2012).
3. Huppert, J.L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33**, 2908–2916 (2005).
4. Bedrat, A., Lacroix, L. & Mergny, J.L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **44**, 1746–1759 (2016).
5. Paeschke, K. *et al.* Telomerase recruitment by the telomere end binding protein-beta facilitates G-quadruplex DNA unfolding in ciliates. *Nat. Struct. Mol. Biol.* **15**, 598–604 (2008).
6. 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).
7. Paeschke, K. *et al.* Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature* **497**, 458–462 (2013).
8. Vannier, J.B., Pavicic-Kaltenbrunner, V., Petalcorin, M.I.R., Ding, H. & Boulton, S.J. RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. *Cell* **149**, 795–806 (2012).
9. Sarkies, P., Reams, C., Simpson, L.J. & Sale, J.E. Epigenetic instability due to defective replication of structured DNA. *Mol. Cell* **40**, 703–713 (2010).
10. 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).
11. Rodriguez, R. *et al.* Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.* **8**, 301–310 (2012).
12. Mendoza, O., Bourdoncle, A., Boulé, J.B., Brosh, R.M. & Mergny, J.L. G-quadruplexes and helicases. *Nucleic Acids Res.* **44**, 1989–2006 (2016).
13. Maizels, N. & Gray, L.T. The G4 genome. *PLoS Genet.* **9**, e1003468 (2013).
14. Rhodes, D. & Lipps, H.J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **43**, 8627–8637 (2015).
15. Hänsel-Hertsch, R., Di Antonio, M. & Balasubramanian, S. DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.* **18**, 279–284 (2017).

16. Schaffitzel, C. *et al.* *In vitro* generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei. *Proc. Natl. Acad. Sci. USA* **98**, 8572–8577 (2001).
17. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **5**, 182–186 (2013).
18. 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).
19. Henderson, A. *et al.* Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.* **42**, 860–869 (2014).
20. 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).
21. Neidle, S. Quadruplex nucleic acids as novel therapeutic targets. *J. Med. Chem.* **59**, 5987–6011 (2016).
22. Chambers, V.S. *et al.* High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* **33**, 877–881 (2015).
23. Mukundan, V.T. Bulges in G4 quadruplexes: broadening the definition of G4 quadruplex-forming sequences. *J. Am. Chem. Soc.* **135**, 5017–5028 (2013).
24. Hänsel-Hertsch, R. *et al.* G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **48**, 1267–1272 (2016).
25. Lam, E.Y.N., Beraldi, D., Tannahill, D. & Balasubramanian, S. G-quadruplex structures are stable and detectable in human genomic DNA. *Nat. Commun.* **4**, 1796 (2013).
26. Xu, H. *et al.* CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nat. Commun.* **8**, 14432 (2017).
27. 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).
28. Kanoh, Y. *et al.* Rif1 binds to G quadruplexes and suppresses replication over long distances. *Nat. Struct. Mol. Biol.* **22**, 889–897 (2015).
29. 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).
30. Liu, H. *et al.* Conformation selective antibody enables genome profiling and leads to discovery of parallel G-quadruplex in human telomeres. *Cell Chem. Biol.* **23**, 1261–1270 (2016).
31. Furey, T.S. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat. Rev. Genet.* **13**, 840–852 (2012).
32. Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I. & Chédin, F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* **45**, 814–825 (2012).
33. Adey, A. *et al.* Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density *in vitro* transposition. *Genome Biol.* **11**, R119 (2010).
34. Teytelman, L. *et al.* Impact of chromatin structures on DNA processing for genomic analyses. *PLoS One* **4**, e6700 (2009).
35. Meyer, C.A. & Liu, X.S. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nat. Rev. Genet.* **15**, 709–721 (2014).
36. Rhee, H.S. & Pugh, B.F. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408–1419 (2011).
37. Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* **33**, 1165–1172 (2015).
38. Schmidl, C., Rendeiro, A.F., Sheffield, N.C. & Bock, C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat. Methods* **12**, 963–5 (2015).
39. Machanick, P. & Bailey, T.L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).
40. Grant, C.E., Bailey, T.L. & Noble, W.S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
41. Schmidt, D. *et al.* ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods* **48**, 240–248 (2009).
42. Nakato, R. & Shirahige, K. Recent advances in ChIP-seq analysis: from quality management to whole-genome annotation. *Brief Bioinform.* **18**, 279–290 (2016).
43. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).
44. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at *arXiv*, https://arxiv.org/abs/1303.3997 (2013).
45. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
46. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
47. Kent, W.J. *et al.* The Human Genome Browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
48. Robinson, J.T. *et al.* Integrative Genome Viewer. *Nat. Biotechnol.* **29**, 24–6 (2011).
49. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
50. 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).
51. Arrigoni, L. *et al.* Standardizing chromatin research: a simple and universal method for ChIP-seq. *Nucleic Acids Res.* **44**, e67 (2016).