

Untargeted CUT&Tag and BG4 CUT&Tag are both enriched at G-quadruplexes and accessible chromatin

AUTHORS

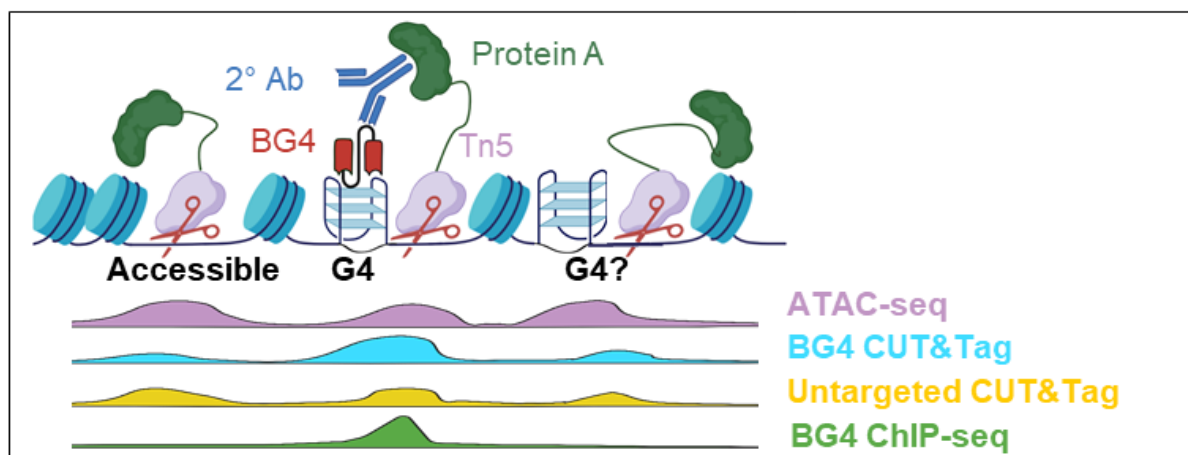
Matthew Thompson¹, Alicia Byrd^{1,2,*}

¹ Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, AR, 72205, USA

² Winthrop P. Rockefeller Cancer Institute, Little Rock, AR, 72205, USA

* To whom correspondence should be addressed. Tel: 601-526-6488; Fax: 1-501-686-8169; Email: akbyrd@uams.edu

GRAPHICAL ABSTRACT



ABSTRACT

G-quadruplex DNA structures (G4s) form within single-stranded DNA in nucleosome-free chromatin. As G4s modulate gene expression and genomic stability, genome-wide mapping of G4s has generated strong research interest. Recently, the Cleavage Under Targets and Tagmentation (CUT&Tag) method was performed with the G4-specific BG4 antibody to target Tn5 transposase to G4s. While this method generated a novel high-resolution map of G4s, we unexpectedly observed a strong correlation between the genome-wide signal distribution of BG4 CUT&Tag and accessible chromatin. To examine whether untargeted Tn5 cutting at accessible chromatin contributes to BG4 CUT&Tag signal, we examined the genome-wide distribution of signal from untargeted (i.e. negative control) CUT&Tag datasets. We observed that untargeted CUT&Tag signal distribution was highly similar to both that of accessible chromatin and of BG4 CUT&Tag. We also observed that BG4 CUT&Tag signal increased at mapped G4s, but this increase was accompanied by a

concomitant increase in untargeted CUT&Tag at the same loci. Consequently, enrichment of BG4 CUT&Tag over untargeted CUT&Tag was not increased at mapped G4s. These results imply that either the vast majority of accessible chromatin regions contain mappable G4s or that the presence of G4s within accessible chromatin cannot reliably be determined using BG4 CUT&Tag alone.

INTRODUCTION

Formation of non-B DNA secondary structures is both a driver and a consequence of myriad biological processes (reviewed in (1)). G-quadruplex DNA structures (G4s) are present throughout regulatory regions of the human genome (2) and influence a diverse array of cellular processes such as gene expression (3–5), DNA replication progression (6–8), and DNA replication origin firing (9, 10). G4s are formed by runs of multiple guanine bases that interact via Hoogsteen base pairing to form planar tetrads (11) that stack into a G4 structure (12). G4s can subsequently be bound and/or resolved by endogenous proteins to modulate transcription and safeguard genomic stability (reviewed in (13, 14)). Additionally, there exist ongoing efforts to selectively stabilize/destabilize G4s *in cellulo* with chemical ligands to modulate translationally relevant biological processes (reviewed in (15)). Consequently, efforts to identify and map folded G4s throughout the human genome has generated significant research interest (reviewed in (16)).

Initial G4 mapping efforts relied on algorithmic prediction (17–19) and next-generation sequencing methods (20) to identify sequences with the potential to form G4s *in vitro*. However, development of the BG4 single-chain antibody that binds specifically to G4s (21) allowed for genome-wide G4 mapping *in cellulo* using chromatin immunoprecipitation with sequencing (ChIP-seq) (2). Recently, new methods of mapping G4s have been created such as G4Access, which uses controlled micrococcal nuclease (MNase) digestion and size selection to isolate folded G4s within subnucleosomal genomic DNA fragments (22). Separate efforts to improve the relatively poor signal-to-noise ratio of ChIP-seq for a variety of genomic targets resulted in the creation of another MNase-based method known as Cleavage Under Targets and Release Using Nuclease (CUT&RUN) (23). Further advancement of CUT&RUN then resulted in the creation of the Tn5 transposase-based Cleavage Under Targets and Tagmentation (CUT&Tag) (24, 25) that reduced both the cost and effort of the method when compared to CUT&RUN. CUT&Tag utilizes a fusion protein comprised of the

antibody-targeting Protein A (and sometimes Protein G) linked to the Tn5 transposase to selectively fragment and tag (i.e. “tagment”) genomic DNA in the vicinity of an antibody target (24). This antibody-based targeting of Tn5 in CUT&Tag contrasts with the lack of targeting of Tn5 tagmentation in the assay for transposase-accessible chromatin with sequencing (ATAC-seq) that maps nucleosome-depleted accessible chromatin (26).

Recently, CUT&Tag has been adapted to map G4s in bulk and single-cell populations using the BG4 single-chain antibody that binds specifically to G4s (27). The use of BG4 CUT&Tag has identified novel G4s that were not previously observed using BG4 ChIP-seq (27), but this is not wholly unexpected, given the improved signal-to-noise ratio of CUT&Tag compared to ChIP-seq (24). However, use of the Tn5 transposase during CUT&Tag creates a susceptibility to off-target DNA cleavage at open chromatin sites that lack the target of interest (24, 28). This could result in CUT&Tag signal accumulation at Tn5-preferred sites under untargeted conditions (i.e. using non-targeting IgG or omitting IgG).

Recent literature has highlighted the utility of identifying regions of the genome to which an abundance of untargeted sequencing reads map. The genome-wide distributions of untargeted signal enrichment in ChIP-seq data (29) and in CUT&RUN data (30) have been generated, and these distributions reflect genomic assembly artefacts and highlight biases intrinsic to the differing methodologies (i.e. DNA shearing via sonication in ChIP-seq and digestion with MNase in CUT&RUN). To our knowledge, no similar map of problematic regions of the genome related to CUT&Tag analyses has been generated.

Given that G4s require an open chromatin environment to form, we hypothesized that Tn5-preferred cleavage at open chromatin during CUT&Tag could generate local CUT&Tag signal enrichment in the vicinity of G4s. To test this, we compiled untargeted CUT&Tag datasets and assessed the local sequence and chromatin environment at sites of signal enrichment in untargeted CUT&Tag. We additionally compared enrichment of BG4 CUT&Tag over untargeted CUT&Tag at mapped G4s in multiple cell lines to assess the capability of BG4 CUT&Tag to specify the presence or absence of G4s within open chromatin, despite the presence of untargeted CUT&Tag signal enrichment at these sites. We observed that although the overall amplitude of BG4 CUT&Tag signal increases at mapped G4s, the enrichment of BG4 CUT&Tag above untargeted CUT&Tag is not increased in the presence of

a mapped G4 within regions of accessible chromatin. These results suggest that reliance solely on peaks within BG4 CUT&Tag data to map G4s can lead to an increase in false negative and/or false positive identification of G4s throughout the genome.

MATERIAL AND METHODS

Read mapping and peak calling

Untargeted CUT&Tag datasets in the Gene Expression Omnibus (GEO) (31) were queried using the following: ((CUT AND Tag AND "homo sapiens")) AND ".bw" AND (("gse"[Filter] OR "gds"[Filter] OR "gsm"[Filter]) AND "Homo sapiens"[Organism]). Untargeted CUT&Tag reads (**Supplementary Table 1**) and reads from other data sources (**Supplementary Tables 2 and 3**) were downloaded from the Sequence Read Archive (SRA) (11) as .fastq files using fastq-dump (<https://github.com/ncbi/sra-tools>) or were downloaded from ENCODE (29) as .bam files pre-mapped to the GRCh38 assembly of the human genome (33). Raw reads (.fastq) were filtered (fastqc), trimmed (34), and aligned (35) to the GRCh38 assembly of the human genome using the Nextflow (36, 37) cutandrun pipeline, v3.2.1 (38). Read counts were normalized by read depth with counts-per-million reads (CPM) normalization with a bin size of 50, and peaks were called with MACS2 (39) (default settings) or with SEACR (40) with a threshold of 0.02. For ENCODE-sourced reads (**Supplementary Table 3**) already mapped to GRCh38, reads were indexed, sorted, and normalized by CPM using deeptools (41) bamCoverage with an effective genome size of 2913022398 and a bin size of 50. When available, peaks of signal enrichment that were called and previously analysed in the source publication were utilized. Unless otherwise indicated, ATAC-seq, CUT&Tag, ChIP-seq, and G4Access datasets were generated in K562 cells (**Supplementary Tables 2 and 3**).

Peak overlap analysis

Genomic intervals present in peaks between more than two datasets were identified using bedtools (42) multiinter, whereas shared or unshared peaks from two datasets were identified using GenomicRanges (43). Read counts at individual peaks were quantified using the UCSC Genome Browser (44) utility bigWigAverageOverBed by quantifying the average normalized read count signal over each base within a peak with non-covered bases being quantified with a signal of zero.

Peak annotation and motif finding

Peaks in untargeted CUT&Tag datasets were annotated using ChIPseeker (45) `annotatePeak` or with HOMER (46) `annotatePeaks` with default settings. Motifs were found with HOMER `findMotifsGenome` with a motif size of 20.

Peak visualization

Heatmaps and metaplots of normalized read counts centred at peaks were generated using `deeptools` (41) `computeMatrix` and `plotHeatmap`. Normalized read counts were visualized across the genome using IGV (47).

RESULTS

Untargeted CUT&Tag samples have a characteristic pattern of genome-wide enrichment

We acquired untargeted CUT&Tag datasets from publicly available data (**Table 1**) to assess whether these data display any pattern of consistent genome-wide enrichment. We selected datasets originating from a variety of cell lines and labs to capture dataset- and cell line-agnostic signal enrichment. Raw reads were reprocessed using `nextflow cutandrun` (38), and peaks were called using both MACS2 (39) (default settings) and SEACR (40) (top 2% of peaks). The high threshold for SEACR was utilized to capture a similar number of peaks as with MACS2 in order to capture genomic regions with read counts which were globally high (SEACR) and/or locally distinguishable (MACS2). We defined consistent signal enrichment as any called peak from either peak caller which was present in more than 33.3% of datasets (at least 7 of 21; similar to ref (30)). 5,163 regions of consistent signal enrichment were identified (**Figure 1A**). This set of conserved peaks is subsequently referred to as untargeted CUT&Tag peaks. The majority of peaks called are shared between both peak calling algorithms (**Supplementary Figure 1**), indicating that untargeted CUT&Tag peaks are frequently both high in amplitude and enriched above local background.

Untargeted CUT&Tag peaks have a unique genome-wide distribution compared to regions of signal enrichment from untargeted ChIP-seq data (48) and untargeted CUT&RUN data (30) (**Figure 1B**). The high proportion of genomic regions of untargeted CUT&RUN signal

enrichment that overlap untargeted ChIP-seq enrichment (80.46%) has been determined to originate from artefactual mapping of short reads to the GRCh38 assembly, where the remaining unshared peaks of untargeted CUT&RUN enrichment (16.21%) were hypothesized to originate from biases in MNase-catalyzed DNA cleavage (30). Similarly, the abundance of unique peaks identified in untargeted ChIP-seq data that do not overlap untargeted CUT&RUN or CUT&Tag peaks (82.55%) are hypothesized to reflect biases related to crosslinking and sonication. Following the same logic, the high percentage of unique peaks (98.80%) in untargeted CUT&Tag data and the low percentage of peaks from untargeted CUT&RUN (1.57%) and ChIP-seq (3.33%) datasets that overlap untargeted CUT&Tag peaks suggest that the observed signal enrichment may be due to DNA sequence or DNA structural biases in Tn5-catalyzed cleavage. This suggested to us that the genome-wide distribution of untargeted CUT&Tag enrichment could resemble the Tn5-based ATAC-seq method which maps accessible chromatin (26). Indeed, untargeted CUT&Tag peaks frequently overlapped sites of ATAC-seq enrichment in K562 cells (**Figure 1C**), suggesting an enrichment at accessible chromatin and regulatory regions.

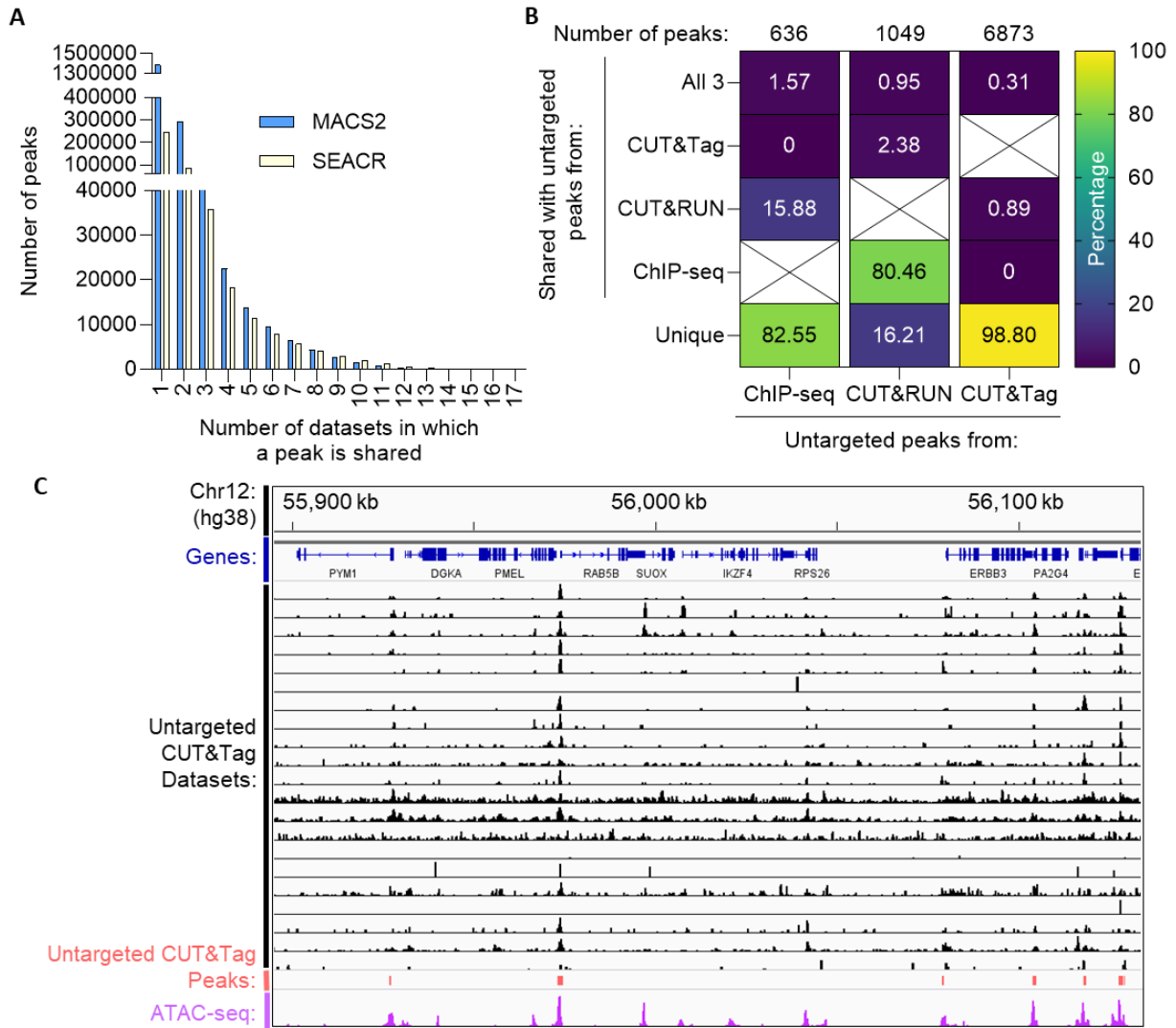


Figure 1. Untargeted CUT&Tag has a consistent genome-wide enrichment. **(A)** Histogram of the number of peaks called by either MACS2 or SEACR that are shared by the specified number of datasets. **(B)** Percentage of unique or shared peaks between peaks of enrichment in untargeted ChIP-seq (48), untargeted CUT&RUN (30), and untargeted CUT&Tag samples. **(C)** Read counts for untargeted CUT&Tag datasets at a representative locus on chromosome 12 overlaid onto the untargeted CUT&Tag consensus peaks and ATAC-seq signal (average of 3 biological replicates) from K562 cells.

Untargeted CUT&Tag peaks are enriched at Tn5-accessible regulatory regions

To quantify the overlap of untargeted CUT&Tag peaks with ATAC-seq-mapped regulatory regions, we first annotated the set of untargeted CUT&Tag consensus peaks. A high percentage (93.94%) of untargeted CUT&Tag peaks are located within 3 kb of a promoter,

with 91.13% of peaks being located less than 1 kb from a promoter (**Figure 2A**). A separate annotation corroborated this, as 79.68% of peaks are annotated as being at a regulatory region, either at a promoter (68.41%) or CpG island (11.27%; **Figure 2B**). Additionally, quantification of signal from ATAC-seq from a representative ENCODE Consortium dataset revealed that ATAC-seq signal increases within 3 kb of a majority of untargeted CUT&Tag peaks (**Figure 2C**). A similar pattern of enrichment was not observed within 3 kb of untargeted ChIP-seq or CUT&RUN peaks (**Figure 2D-E**). This demonstrates that untargeted CUT&Tag peaks colocalize with Tn5-accessible chromatin within regulatory regions.

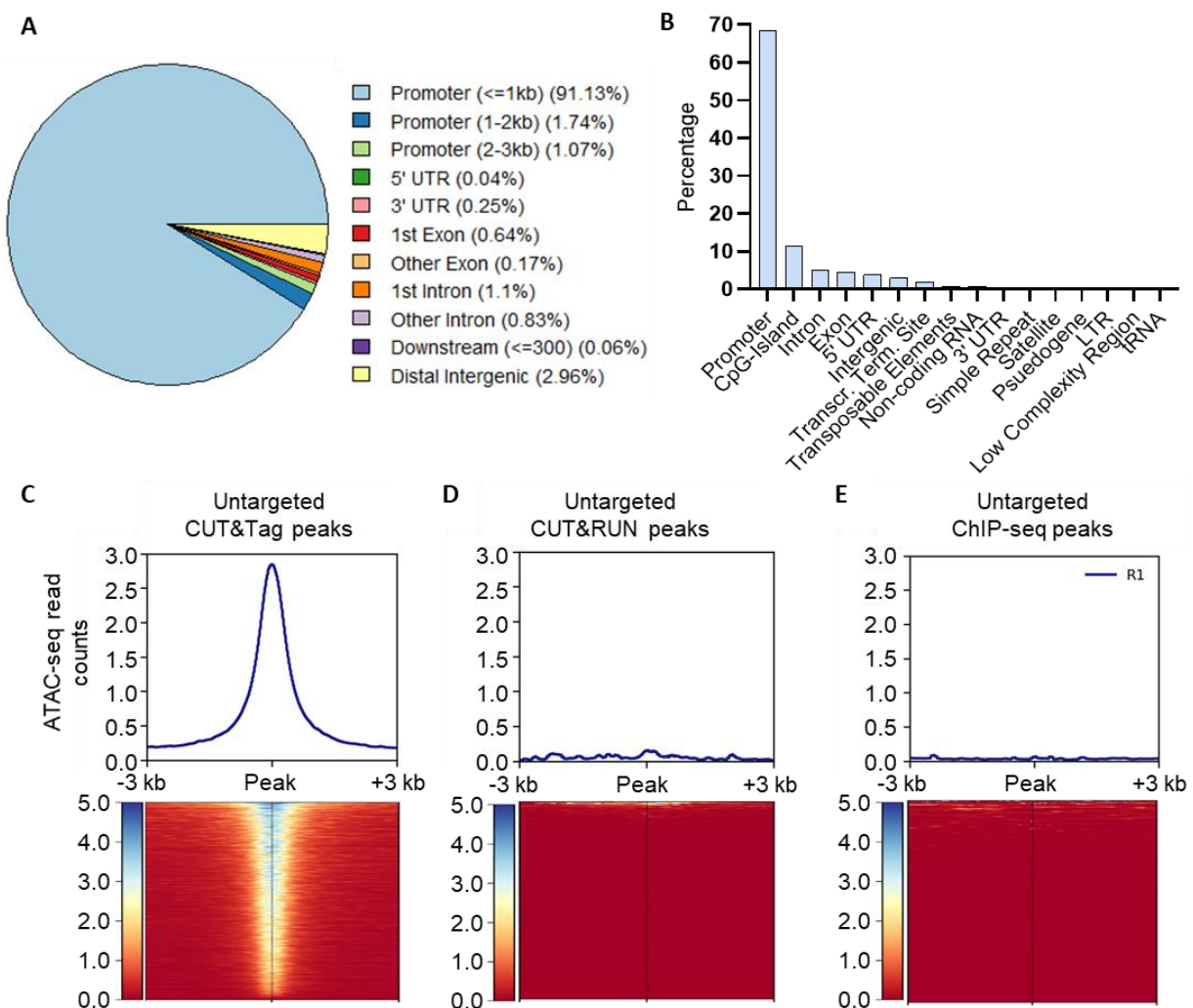


Figure 2. Untargeted CUT&Tag peaks colocalize with regulatory regions and accessible chromatin. **(A)** Proportions of annotations overlapping with untargeted CUT&Tag peaks using ChIPseeker (45). **(B)** Proportions of annotations overlapping with untargeted CUT&Tag peaks using HOMER (46). Metaplots of ATAC-seq signal (ENCSR868FGK; average of 3

biological replicates) at peaks of signal enrichment in untargeted CUT&Tag (**C**), CUT&RUN (30) (**D**), and ChIP-seq (48) (**E**) samples.

Untargeted CUT&Tag peaks colocalize with G4s

Given that accessible regulatory regions of the human genome have the propensity to form DNA secondary structures such as G4s (2), we wondered whether CUT&Tag problematic regions colocalized with G4s. Identification of known sequence motifs that occur in untargeted CUT&Tag peaks revealed the prevalence of motifs with G- and C-rich sequences with tandem runs of guanines or cytosines (**Figure 3A**). Additionally, the untargeted CUT&Tag peaks contain known binding motifs for the Sp1 transcription factor, a *bone fide* G4-binding protein (3, 4). *De novo* motifs also include Sp1 motifs, motifs for the YY1 G4-binding transcription factor (49), and tandem runs of guanines and cytosines (**Figure 3B**). *De novo* motifs also contain TTGxxG/C motifs, which have been previously identified at Tn5 insertion sites (50). The proportion of untargeted CUT&Tag peaks that overlap a G4 was calculated for each set of untargeted CUT&Tag peaks that are shared by 1 or more datasets. The least reproducible untargeted CUT&Tag peaks (present in 4 or fewer datasets) have a low proportion of peaks that overlap a mapped G4 (**Figure 3C**). However, as peak reproducibility increases, the proportion of untargeted CUT&Tag peaks that overlap a mapped G4 also increases, reaching saturation when peaks are shared by at least 5 datasets (**Figure 3C**). This indicates that untargeted CUT&Tag signal is reproducibly enriched at G4s that are algorithmically predicted (pqsfinder) (51), able to form *in vitro* (G4-seq) (52), and observed in human cell lines through use of G4Access (22), BG4 CUT&Tag (27), and/or BG4 ChIP-seq (53).

Additionally, we quantified the degree of colocalization between the set of untargeted CUT&Tag peaks (present in at least 7 of 21 datasets) with two different BG4-based methods of mapping G4s. The majority (95.91%) of the untargeted CUT&Tag peaks overlap G4s mapped by BG4 ChIP-seq (53) and/or BG4 CUT&Tag (27) (**Figure 3D**). However, 13.96% of untargeted CUT&Tag peaks colocalize with only BG4 CUT&Tag G4s and not with BG4 ChIP-seq G4s.

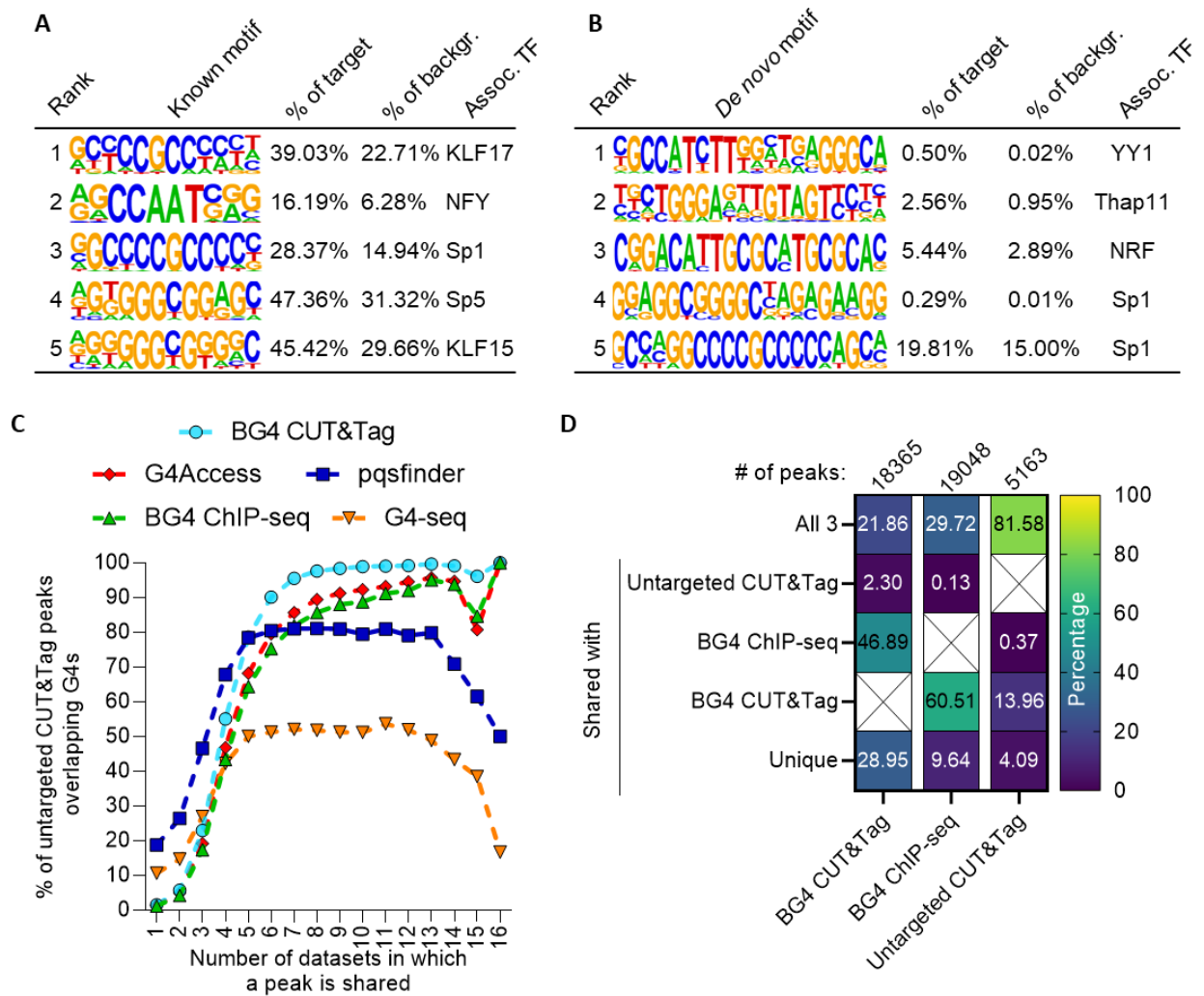


Figure 3. Untargeted CUT&Tag peaks colocalize with mapped G4s. **(A)** Top 5 known motifs and associated transcription factors (TF) present in untargeted CUT&Tag peaks identified by HOMER (46). All p-values $< 1 \times 10^{-125}$ by binomial test. **(B)** Top 5 *de novo* motifs and associated transcription factors (TF) present in untargeted CUT&Tag peaks identified by HOMER (46). All p-values $< 1 \times 10^{-20}$ by binomial test. **(C)** Proportion of untargeted CUT&Tag peaks present in the specified number of datasets that overlap a G4 mapped by BG4 CUT&Tag (27), BG4 ChIP-seq (53), G4Access (22), or mapped *in vitro* by G4-seq (52) or *in silico* by pqsfinder (51). **(D)** Proportion of overlap between untargeted CUT&Tag peaks and BG4 CUT&Tag (27) and/or BG4 ChIP-seq (53).

Enrichment of BG4 CUT&Tag over untargeted CUT&Tag is not increased at G4s

Given our prior observation of colocalized enrichment of both BG4 CUT&Tag and untargeted CUT&Tag, we hypothesized that untargeted Tn5 cutting at accessible chromatin could contribute to signal enrichment in CUT&Tag regardless of the inclusion of the BG4 antibody. This could result in a confounded measurement of the presence of a folded G4 *in cellulo* when using BG4 CUT&Tag, as untargeted CUT&Tag signal could also be enriched at accessible chromatin in the presence of a mapped G4.

To test this hypothesis, we compared the enrichment of BG4 CUT&Tag signal over untargeted CUT&Tag at mapped G4s (22, 27, 51–53). We reasoned that, if BG4 CUT&Tag signal is derived from BG4 affinity to G4s and is not derived from untargeted cutting at accessible chromatin, then we should observe an increased enrichment of BG4 CUT&Tag above untargeted CUT&Tag signal at G4s that is not observed in the absence of a G4. Therefore, we first compared the distribution of read counts of BG4 CUT&Tag and untargeted CUT&Tag signal at BG4 CUT&Tag peaks (27) that overlap or do not overlap a mapped G4. Both BG4 CUT&Tag and untargeted CUT&Tag read counts (27) increase at mapped G4s, with the strongest increase at G4s that were mapped *in cellulo* by BG4 ChIP-seq (53) (**Figure 4A, Supplementary Figure 2A**) and G4Access (22) (**Figure 4B**). Additionally, both BG4 CUT&Tag and untargeted CUT&Tag read counts increase at G4s mapped *in vitro* with G4-seq (52) (**Figure 4C, Supplementary Figure 2B**) and *in silico* with pqsfinder (51) (**Figure 4D, Supplementary Figure 2C**).

The concomitant increase in signal for both BG4-targeted and untargeted CUT&Tag suggests a possible increase in chromatin accessibility at these sites, as Tn5 cutting at accessible chromatin can occur in targeted or untargeted CUT&Tag reactions, although the signal amplitude due to Tn5 cutting at off-target accessible chromatin is generally low (24). In line with this, an increase in ATAC-seq signal is observed at the previously examined BG4 CUT&Tag peaks that overlap a mapped G4 (**Figure 4E, Supplementary Figure 2D**). Given that G4 formation requires nucleosome-depleted ssDNA, this is expected. However, it is important to ensure that the increase in BG4 CUT&Tag signal at G4s is related to BG4 antibody-based targeting of Tn5 to G4s (21) and does not occur due the local increase in chromatin accessibility.

If BG4 CUT&Tag signal reflects BG4 affinity and not chromatin accessibility, we would expect to observe an increase in BG4 CUT&Tag signal greater in magnitude than that of untargeted CUT&Tag (i.e. increased enrichment over background) at mapped G4s. To measure this, we quantified the ratio of BG4 CUT&Tag signal over untargeted CUT&Tag signal at each of the individual CUT&Tag peaks that overlap or do not overlap a mapped G4. This reveals a median two-fold enrichment of BG4 CUT&Tag read counts above untargeted CUT&Tag read counts at BG4 CUT&Tag peaks that do not overlap a mapped G4. Unexpectedly, this two-fold ratio of BG4-targeted CUT&Tag over untargeted CUT&Tag does not appreciably change in the presence of a mapped G4 (**Figure 4F, Supplementary Figure 2E**). In fact, for BG4 ChIP-seq (53), G4-seq (52), and pqsfinder (51), the median ratio of BG4 CUT&Tag signal over untargeted CUT&Tag signal at individual CUT&Tag peaks is greater at regions that do not overlap a mapped G4 than those that do overlap a mapped G4. This suggests that CUT&Tag reads from untargeted Tn5 cutting at accessible chromatin strongly contribute to BG4 CUT&Tag signal in addition to the contribution of CUT&Tag reads from BG4-targeted Tn5 cutting at folded G4s *in cellulo*. Supporting this suggestion, CUT&Tag signal enrichment for both BG4 CUT&Tag and untargeted CUT&Tag can be observed at ATAC-seq sites in the absence of a visible enrichment in BG4 ChIP-seq signal (**Figure 4G, Supplementary Figure 2F**). G4Access signal is also observed at these sites, but we also observed a subset of these sites which lack overlap with either a G4-seq G4 (**Figure 4G**) or both G4-seq and pqsfinder G4s (**Supplementary Figure 2F**). Two possible explanations for these results are: BG4 CUT&Tag and G4Access have increased resolution over BG4 ChIP-seq, giving them signal enrichment at G4s that BG4 ChIP-seq is unable to detect, or CUT&Tag and G4Access both share a similar background resulting from enzymatic (Tn5 or MNase) cutting at accessible chromatin. To test these possibilities, we performed an analysis of CUT&Tag signal enrichment over background at accessible chromatin sites in the presence or absence of a mapped G4.

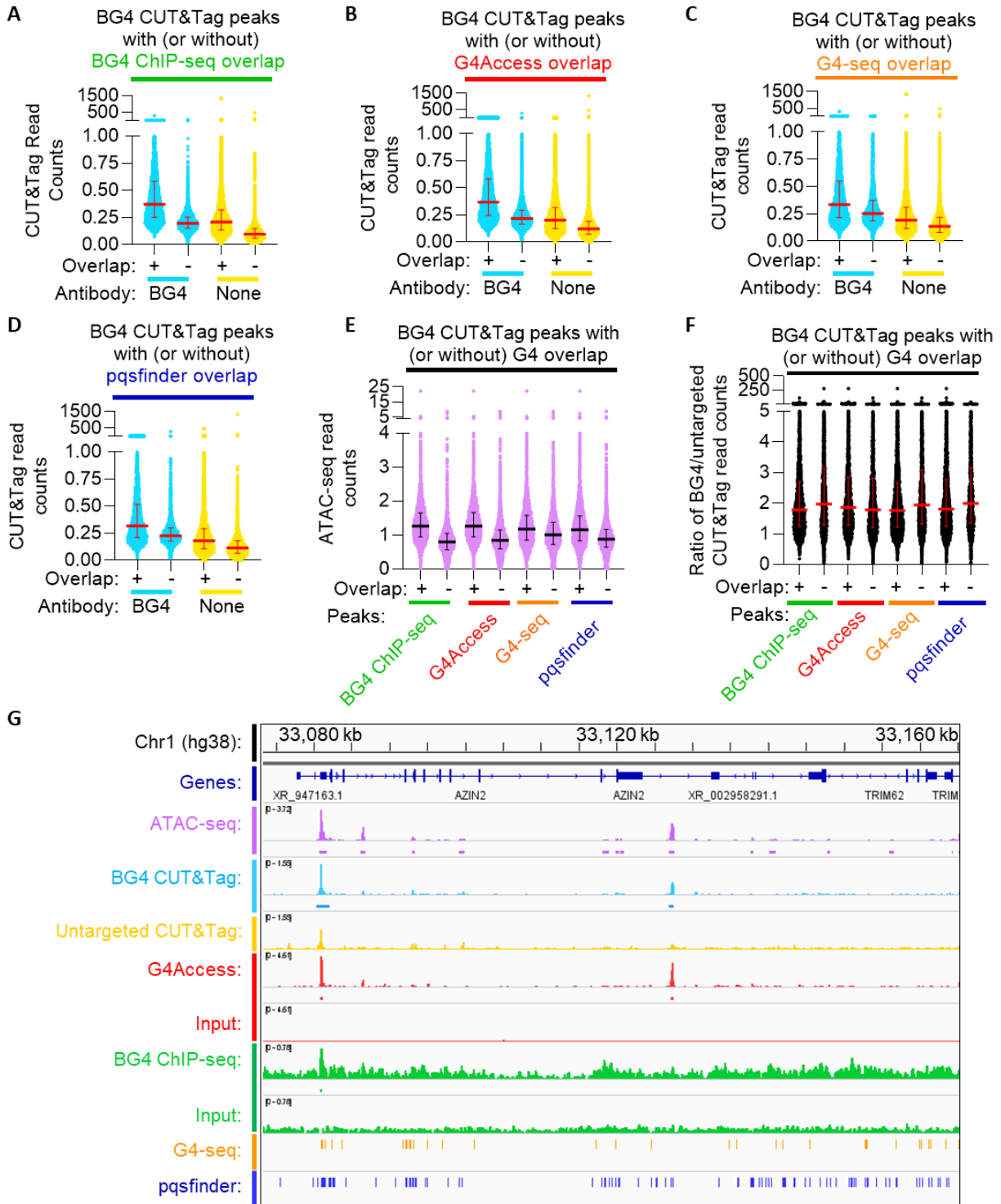


Figure 4. Enrichment of BG4 CUT&Tag over untargeted CUT&Tag does not increase at G4s in K562 cells. **(A-D)** BG4 or untargeted CUT&Tag signal (27) read counts were quantified at BG4 CUT&Tag peaks that overlap or do not overlap a G4 mapped by BG4 ChIP-seq (53) **(A)**, G4Access (22) **(B)**, G4-seq (52) **(C)**, or pqsfinder (51) **(D)** in K562 cells. Data are the average of 3 biological replicates, and the median and interquartile ranges are plotted in red. **(E)** ATAC-

seq read counts at BG4 CUT&Tag peaks that overlap or do not overlap a G4 mapped by the indicated method are plotted. Data are the average of 3 biological replicates, and the median and interquartile ranges are plotted in black. **(F)** Ratio of read counts of BG4 CUT&Tag (27) and untargeted CUT&Tag (27) quantified at individual CUT&Tag peaks that overlap or do not overlap a G4 mapped by the indicated method. Data are the average of 3 biological replicates, and the median and interquartile ranges are plotted in red. **(G)** Read counts (signal traces) and called peaks (rectangles) of ATAC-seq, BG4 CUT&Tag (27), untargeted CUT&Tag (27), G4Access (22), G4Access input (22), BG4 ChIP-seq (53) and ChIP-seq input (53) from K562 cells with G4-seq (52) and pqsfinder-predicted G4s (51) at a representative locus on chromosome 1. ATAC-seq, BG4 CUT&Tag, and untargeted CUT&Tag reads are the average of 3 biological replicates, G4Access reads are the average of 2 biological replicates, and G4Access input, BG4 ChIP-seq, and ChIP-seq input are data from 1 biological replicate. BG4 ChIP-seq peaks are the average of two biological replicates.

Enrichment of BG4 CUT&Tag over untargeted CUT&Tag is not increased at G4s within accessible chromatin

To investigate whether CUT&Tag data is enriched at antibody targets compared to all accessible chromatin, we examined enrichment of target CUT&Tag data over untargeted CUT&Tag data at ATAC-seq peaks that overlap or do not overlap ChIP-seq peaks of the CUT&Tag target. Using histone post-translational modifications (PTMs) as controls, we observed a dramatic increase in CUT&Tag read counts (24) at ATAC-seq peaks that overlap the H3K4me3 or H3K27me3 ChIP-seq peaks compared to ATAC-seq peaks that do not overlap a ChIP-seq peak (**Supplementary Figure 3A**). The CUT&Tag read counts at ATAC-seq peaks that do not overlap the appropriate histone PTM ChIP-seq peaks are close to zero for both H3K4me3 and H3K27me3 (**Supplementary Figure 3A**). Additionally, enrichment of target CUT&Tag signal compared to untargeted CUT&Tag is observed at ATAC-seq peaks that overlap the appropriate ChIP-seq peak, as opposed to a lack of enrichment at ATAC-seq peaks which do not overlap a ChIP-seq peak (**Supplementary Figure 3B**). These data demonstrate that for histone PTMs, targeted CUT&Tag signal is significantly enriched above untargeted CUT&Tag at the expected ChIP-seq mapped sites and likely reflects antibody-targeted Tn5 cleavage.

To investigate whether BG4 CUT&Tag signal is enriched over untargeted CUT&Tag at G4s specifically and not at all accessible chromatin, the same analysis was performed. Read counts of BG4 CUT&Tag (27) and untargeted CUT&Tag (27) were examined at ATAC-seq sites that overlap a mapped G4 (22, 51–53). Both BG4 CUT&Tag and untargeted CUT&Tag reads increase at ATAC-seq peaks that overlap a mapped G4 (**Figure 5A, Supplementary Figure 4A**). However, the enrichment of BG4 CUT&Tag over untargeted CUT&Tag does not appreciably increase at mapped G4s (**Figure 5B, Supplementary Figure 4B**). These results suggest that enrichment of BG4 CUT&Tag signal over untargeted CUT&Tag does not necessarily signal the presence of a G4 within accessible chromatin. Indeed, we again observed enrichment of BG4 CUT&Tag (27), untargeted CUT&Tag (27), and G4Access (22) at accessible chromatin in the absence of an accompanying BG4 ChIP-seq mapped G4, but this enrichment was not observed for CUT&Tag reactions targeting histone PTMs (**Figure 5C**). In total, these results suggest that the genome-wide distribution of reads generated by BG4 CUT&Tag and untargeted CUT&Tag are strongly influenced by susceptibility of the local chromatin environment to untargeted Tn5 cutting.

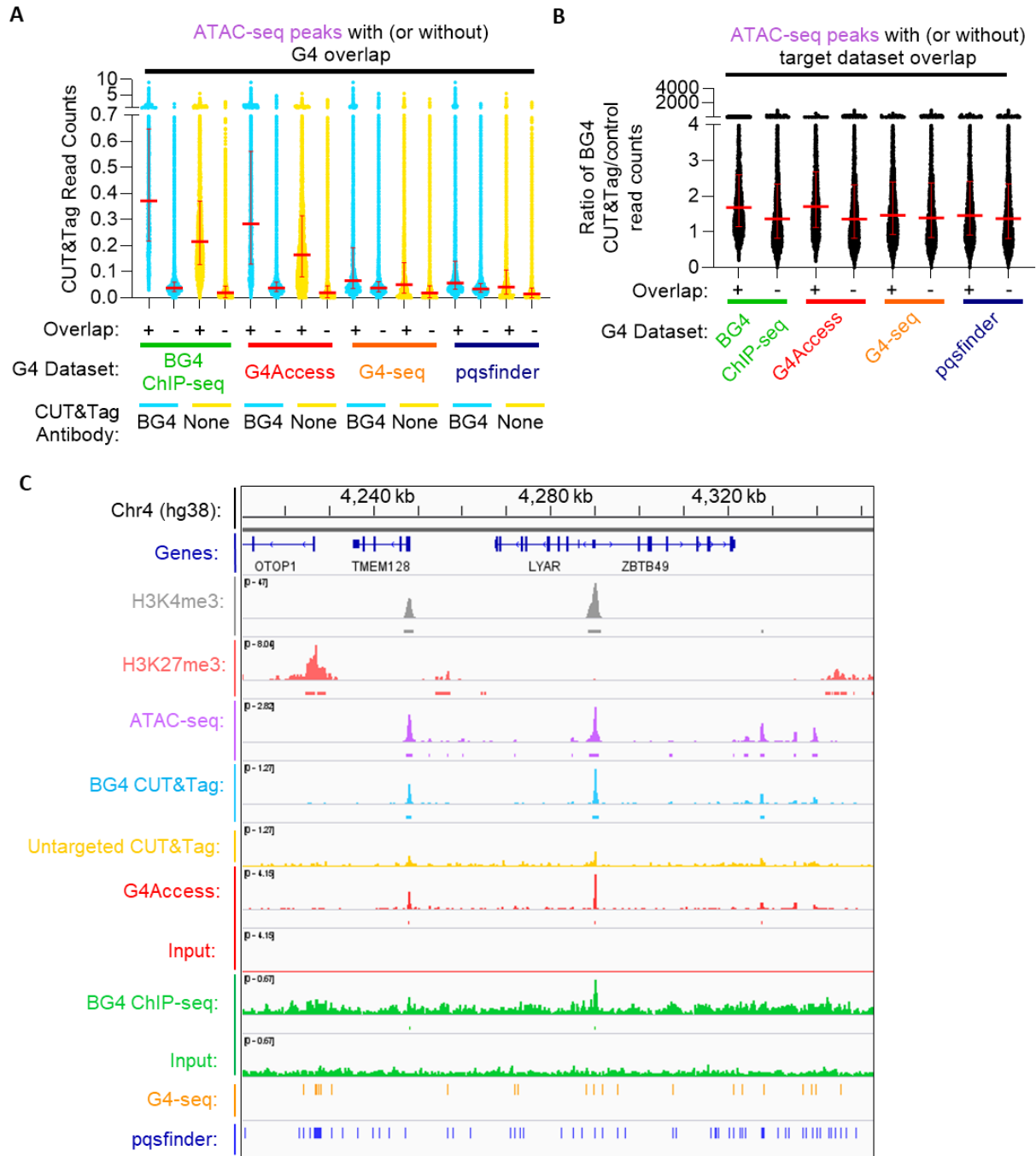


Figure 5. Enrichment of BG4 CUT&Tag over untargeted CUT&Tag does not increase at G4s within accessible chromatin in K562 cells. **(A)** Read counts of BG4 CUT&Tag (27) and untargeted CUT&Tag (27) from K562 cells at ATAC-seq peaks that overlap or do not overlap a mapped G4 (22, 51–53). Data are the average of 3 biological replicates, and the median and interquartile ranges are plotted in red. **(B)** Ratio of read counts of BG4 CUT&Tag (27) and untargeted CUT&Tag (27) from K562 cells quantified at ATAC-seq peaks that overlap or do not overlap a G4 mapped by the indicated method (22, 51–53). Data are the average of 3

biological replicates, and the median and interquartile ranges are plotted in red. (C) Read counts (signal traces) and called peaks (rectangles) of H3K4me3 CUT&Tag (24), H3K27me3 CUT&Tag (24), ATAC-seq, BG4 CUT&Tag (27), untargeted CUT&Tag (27), G4Access (22), G4Access input (22), BG4 ChIP-seq (53), and ChIP-seq input (53) from K562 cells with G4-seq (52) and pqsfinder-predicted G4s (51) at a representative locus on chromosome 4. Read counts of H3K4me3 CUT&Tag, H3K27me3 CUT&Tag, G4Access input, BG4 ChIP-seq, and ChIP-seq input are from 1 biological replicate. Read counts from G4Access are the average of 2 biological replicates. Read counts from ATAC-seq, BG4 CUT&Tag, and untargeted CUT&Tag are the average of 3 biological replicates. BG4 ChIP-seq peaks are the average of two biological replicates.

DISCUSSION

The need to predict and map G4s (reviewed in (16)) has spurred technical innovation, and a diverse suite of G4 mapping tools exists to meet multiple experimental needs. For example, full-length antibodies (54), single-chain antibodies (21), nanobodies (55), protein-based probes (56), and G4-ligand-based probes (57) have been utilized for chromatin immunoprecipitation- (58) and *in situ* Tn5-based sequence mapping methods (27, 57). Additionally, the G4Access method uses the G-rich sequence cutting preference of MNase combined with the inhibition of MNase cutting by nucleosomes and folded G4s to map G4s within accessible chromatin (22). While each method has limitations, BG4 CUT&Tag (27) has been noted for its high resolution compared to BG4 ChIP-seq and low input requirements (24) which can enable G4 mapping of single cell populations (27) and precious patient-derived samples. However, CUT&Tag analyses can be influenced by untargeted Tn5 cutting at accessible chromatin (24, 28). Indeed, simply by reducing the ionic strength of the buffer during tagmentation, Tn5-based cleavage can be used to map accessible chromatin in the vicinity of antibody targets, rather than solely mapping the site of antibody binding (59). Given the dynamic nature of G4s and their high degree of overlap with accessible regulatory chromatin, it is important to characterize which G4 mapping methods can be used to distinguish the presence or absence of G4s without mapping accessible chromatin environments instead.

The initially-reported method to distinguish between antibody-targeted and untargeted Tn5 tagmentation at accessible chromatin for CUT&Tag analyses used read count-based discrimination to separate high read count targeted signals from low read count untargeted signals (24). The read count difference between targeted CUT&Tag of histone PTMs compared to untargeted Tn5 tagmentation is sufficient to discriminate between on- and off-target CUT&Tag signal enrichment (**Supplementary Figure 3A**). For peaks of BG4 CUT&Tag signal with the highest read count enrichment, this method could be employed to identify high-confidence peaks of CUT&Tag enrichment. Indeed, this method has been utilized to call regions with the top 5% of signal using SEACR (40) without a negative control dataset (27). However, the high degree of overlap between the read count distributions of BG4-targeted and untargeted CUT&Tag combined with the simultaneous enrichment of both BG4-targeted and untargeted CUT&Tag signals at mapped G4s (**Figure 4**) would create a large population of false negative or false positive G4s using this method of read count discrimination. Additionally, the simultaneous enrichment of both BG4-targeted and untargeted CUT&Tag reads at mapped G4s and accessible chromatin may increase false negative results (i.e. lack of a peak call) when a negative control is included during peak calling, as an increase in targeted BG4 CUT&Tag signal at a folded G4 could be masked by the simultaneous increased read counts of untargeted CUT&Tag at the same locus.

Consequently, the results of our analysis place limitations on identifying targeted signal enrichment at antibody targets over untargeted enrichment at accessible chromatin using CUT&Tag. This limitation poses minimal issues for CUT&Tag of highly abundant, high-occupancy targets such as histone PTMs, as enrichment of target CUT&Tag signal over untargeted CUT&Tag signal strongly increases at mapped target sites within accessible chromatin (**Supplementary Figure 3**). However, BG4-targeted CUT&Tag for G4s does not result in an appreciable increase in signal enrichment over untargeted signal enrichment at mapped G4s within accessible chromatin (**Figure 5 and Supplementary Figure 4**). Additionally, we observed the presence of BG4 CUT&Tag signal enrichment at accessible chromatin that is not readily observed using BG4 ChIP-seq (53) (**Figure 5C**).

The lack of a strong increase in BG4 CUT&Tag enrichment over untargeted CUT&Tag at mapped G4s and the identification of novel BG4 peaks of enrichment when compared to BG4 ChIP-seq (53) could be the result of a combination of multiple factors. First, Tn5-based

DNA cleavage methods result in an increase in signal-to-noise when compared to BG4 ChIP-seq, so BG4 CUT&Tag signal enrichment at sites lacking visible BG4 ChIP-seq enrichment is not unexpected. Second, Tn5-based methods can result in untargeted cleavage at accessible chromatin (24, 28), suggesting that at least some of the sites of BG4 CUT&Tag enrichment could be the result of Tn5 cleavage at accessible chromatin. The strong genome-wide correlation between ATAC-seq, G4Access, BG4 CUT&Tag, and untargeted CUT&Tag (**Figure 5C**) suggests that chromatin accessibility strongly dictates enrichment of both BG4-targeted and untargeted CUT&Tag signal. This is also not surprising, as G4s form in ssDNA within nucleosome-depleted chromatin, and G4 formation may actively exclude nearby nucleosomes (5, 60, 61). Third, other literature has implicated both hyperactive Tn5 (62) and Protein A (63) in binding to G4s, suggesting that untargeted cleavage may also occur due to unanticipated G4 binding by transposase components at existing G4s.

When taken as a whole, our findings suggest that stringent biochemical validation of sites of BG4 CUT&Tag enrichment should always be performed to strengthen the confidence in BG4 CUT&Tag analyses, such as knockdown of G4-resolving enzymes followed by an examination of BG4 CUT&Tag signal increase (as in (22)), confirmation of individual G4s with circular dichroism spectroscopy (reviewed in (64)), and/or *in vitro* binding assays examining direct interaction between the examined G4s and other colocalizing proteins. However, as validation of every peak of enrichment across an entire genome is not realistic to perform, this lessens our confidence in BG4 CUT&Tag analyses in the absence of other corroborating data.

DATA AVAILABILITY

All data utilized for this analysis is available from GEO (10), SRA (11), or ENCODE (29) at the accessions included in **Supplementary Tables 1, 2, and 3**.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

AUTHOR CONTRIBUTIONS

MDT: Analysis, Methodology, Writing—original draft. AKB: Conceptualization, Writing—review & editing.

FUNDING

This work was supported by the National Institutes of Health P20GM121293. Funding for open access charge: National Institutes of Health P20GM121293.

CONFLICT OF INTEREST

None declared.

REFERENCES

1. Wang,G. and Vasquez,K.M. (2023) Dynamic alternative DNA structures in biology and disease. *Nat. Rev. Genet.*, **24**, 211–234.
2. Hänsel-Hertsch,R., Beraldi,D., Lensing,S.V., Marsico,G., Zyner,K., Parry,A., Di Antonio,M., Pike,J., Kimura,H., Narita,M., *et al.* (2016) G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.*, **48**, 1267–1272.
3. Raiber,E.-A., Kranaster,R., Lam,E., Nikan,M. and Balasubramanian,S. (2012) A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Res.*, **40**, 1499–1508.
4. Lago,S., Nadai,M., Cernilogar,F.M., Kazerani,M., Domíniguez Moreno,H., Schotta,G. and Richter,S.N. (2021) Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nat. Commun.*, **12**, 3885.
5. Esain-Garcia,I., Kirchner,A., Melidis,L., Tavares,R. de C.A., Dhir,S., Simeone,A., Yu,Z., Madden,S.K., Hermann,R., Tannahill,D., *et al.* (2024) G-quadruplex DNA structure is a positive regulator of MYC transcription. *Proc. Natl. Acad. Sci.*, **121**, e2320240121.
6. Butler,T.J., Estep,K.N., Sommers,J.A., Maul,R.W., Moore,A.Z., Bandinelli,S., Cucca,F., Tuke,M.A., Wood,A.R., Bharti,S.K., *et al.* (2020) Mitochondrial genetic variation is enriched in G-quadruplex regions that stall DNA synthesis in vitro. *Hum. Mol. Genet.*, **29**, 1292–1309.
7. Dahan,D., Tsirkas,I., Dovrat,D., Sparks,M.A., Singh,S.P., Galletto,R. and Aharoni,A. (2018) Pif1 is essential for efficient replisome progression through lagging strand G-quadruplex DNA secondary structures. *Nucleic Acids Res.*, **46**, 11847–11857.
8. Dong,D.W., Pereira,F., Barrett,S.P., Kolesar,J.E., Cao,K., Damas,J., Yatsunyk,L.A., Johnson,F.B. and Kaufman,B.A. (2014) Association of G-quadruplex forming sequences with human mtDNA deletion breakpoints. *BMC Genomics*, **15**, 677.
9. Poulet-Benedetti,J., Tonnerre-Doncarli,C., Valton,A.-L., Laurent,M., Gérard,M., Barinova,N., Parisis,N., Massip,F., Picard,F. and Prioleau,M.-N. (2023) Dimeric G-quadruplex motifs-induced NFRs determine strong replication origins in vertebrates. *Nat. Commun.*, **14**, 4843.
10. Prorok,P., Artufel,M., Aze,A., Coulombe,P., Peiffer,I., Lacroix,L., Guédin,A., Mergny,J.-L., Damaschke,J., Schepers,A., *et al.* (2019) Involvement of G-quadruplex regions in mammalian replication origin activity. *Nat. Commun.*, **10**, 3274.

11. Gellert,M., Lipsett,M.N. and Davies,D.R. (1962) Helix formation by guanylic acid. *Proc. Natl. Acad. Sci.*, **48**, 2013–2018.
12. Williamson,J.R., Raghuraman,M.K. and Cech,T.R. (1989) Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell*, **59**, 871–880.
13. Meier-Stephenson,V. (2022) G4-quadruplex-binding proteins: review and insights into selectivity. *Biophys. Rev.*, **14**, 635–654.
14. Sato,K. and Knipscheer,P. (2023) G-quadruplex resolution: From molecular mechanisms to physiological relevance. *DNA Repair*, **130**, 103552.
15. Figueiredo,J., Mergny,J.-L. and Cruz,C. (2024) G-quadruplex ligands in cancer therapy: Progress, challenges, and clinical perspectives. *Life Sci.*, **340**, 122481.
16. Galli,S., Flint,G., Růžičková,L. and Di Antonio,M. Genome-wide mapping of G-quadruplex DNA: a step-by-step guide to select the most effective method. *RSC Chem. Biol.*, **5**, 426–438.
17. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
18. Kikin,O., D’Antonio,L. and Bagga,P.S. (2006) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.*, **34**, W676–W682.
19. Eddy,J. and Maizels,N. (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.*, **34**, 3887–3896.
20. Chambers,V.S., Marsico,G., Boutell,J.M., Di Antonio,M., Smith,G.P. and Balasubramanian,S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.*, **33**, 877–881.
21. Biffi,G., Tannahill,D., McCafferty,J. and Balasubramanian,S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.
22. Esnault,C., Magat,T., Zine El Aabidine,A., Garcia-Oliver,E., Cucchiari,A., Bouchouika,S., Lleres,D., Goerke,L., Luo,Y., Verga,D., *et al.* (2023) G4access identifies G-quadruplexes and their associations with open chromatin and imprinting control regions. *Nat. Genet.*, **55**, 1359–1369.
23. Skene,P.J. and Henikoff,S. (2017) An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife*, **6**, e21856.

24. Kaya-Okur,H.S., Wu,S.J., Codomo,C.A., Pledger,E.S., Bryson,T.D., Henikoff,J.G., Ahmad,K. and Henikoff,S. (2019) CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.*, **10**, 1930.
25. Kaya-Okur,H.S., Janssens,D.H., Henikoff,J.G., Ahmad,K. and Henikoff,S. (2020) Efficient low-cost chromatin profiling with CUT&Tag. *Nat. Protoc.*, **15**, 3264–3283.
26. Buenrostro,J.D., Giresi,P.G., Zaba,L.C., Chang,H.Y. and Greenleaf,W.J. (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods*, **10**, 1213–1218.
27. Hui,W.W.I., Simeone,A., Zyner,K.G., Tannahill,D. and Balasubramanian,S. (2021) Single-cell mapping of DNA G-quadruplex structures in human cancer cells. *Sci. Rep.*, **11**, 23641.
28. Wang,M. and Zhang,Y. (2021) Tn5 transposase-based epigenomic profiling methods are prone to open chromatin bias. bioRxiv doi: <https://doi.org/10.1101/2021.07.09.451758>, 10 July 2021, pre-print: not peer-reviewed.
29. The ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, **306**, 636–640.
30. Nordin,A., Zambanini,G., Pagella,P. and Cantù,C. (2023) The CUT&RUN suspect list of problematic regions of the genome. *Genome Biol.*, **24**, 185.
31. Edgar,R., Domrachev,M. and Lash,A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.
32. Leinonen,R., Sugawara,H. and Shumway,M. (2011) The Sequence Read Archive. *Nucleic Acids Res.*, **39**, D19–D21.
33. Schneider,V.A., Graves-Lindsay,T., Howe,K., Bouk,N., Chen,H.-C., Kitts,P.A., Murphy,T.D., Pruitt,K.D., Thibaud-Nissen,F., Albracht,D., *et al.* (2017) Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res.*, **27**, 849–864.
34. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10–12.
35. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.

36. Di Tommaso,P., Chatzou,M., Floden,E.W., Barja,P.P., Palumbo,E. and Notredame,C. (2017) Nextflow enables reproducible computational workflows. *Nat. Biotechnol.*, **35**, 316–319.
37. Ewels,P.A., Peltzer,A., Fillinger,S., Patel,H., Alneberg,J., Wilm,A., Garcia,M.U., Di Tommaso,P. and Nahnsen,S. (2020) The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.*, **38**, 276–278.
38. Cheshire,C., charlotte-west, Rönkkö,T., bot, nf-core, Patel,H., tamara-hodgetts, Ladd,D., Thiery,A., Fields,C., Deu-Pons,J., *et al.* (2023) nf-core/cutandrun: nf-core/cutandrun v3.2.1 Mercury Turkey. 10.5281/zenodo.10040206.
39. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W., *et al.* (2008) Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
40. Meers,M.P., Tenenbaum,D. and Henikoff,S. (2019) Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin*, **12**, 42.
41. Ramírez,F., Dündar,F., Diehl,S., Grüning,B.A. and Manke,T. (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.*, **42**, W187–W191.
42. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
43. Lawrence,M., Huber,W., Pagès,H., Aboyoun,P., Carlson,M., Gentleman,R., Morgan,M.T. and Carey,V.J. (2013) Software for Computing and Annotating Genomic Ranges. *PLOS Comput. Biol.*, **9**, e1003118.
44. Kent,W.J., Sugnet,C.W., Furey,T.S., Roskin,K.M., Pringle,T.H., Zahler,A.M. and Haussler, and D. (2002) The Human Genome Browser at UCSC. *Genome Res.*, **12**, 996–1006.
45. Yu,G., Wang,L.-G. and He,Q.-Y. (2015) ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, **31**, 2382–2383.
46. Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H. and Glass,C.K. (2010) Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell*, **38**, 576–589.
47. Thorvaldsdóttir,H., Robinson,J.T. and Mesirov,J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.*, **14**, 178–192.

48. Amemiya,H.M., Kundaje,A. and Boyle,A.P. (2019) The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.*, **9**, 9354.
49. Li,L., Williams,P., Ren,W., Wang,M.Y., Gao,Z., Miao,W., Huang,M., Song,J. and Wang,Y. (2021) YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. *Nat. Chem. Biol.*, **17**, 161–168.
50. Zhang,H., Lu,T., Liu,S., Yang,J., Sun,G., Cheng,T., Xu,J., Chen,F. and Yen,K. (2021) Comprehensive understanding of Tn5 insertion preference improves transcription regulatory element identification. *NAR Genomics Bioinforma.*, **3**, lqab094.
51. Hon,J., Martínek,T., Zendulka,J. and Lexa,M. (2017) pqsfinder: an exhaustive and imperfection-tolerant search tool for potential quadruplex-forming sequences in R. *Bioinformatics*, **33**, 3373–3379.
52. Marsico,G., Chambers,V.S., Sahakyan,A.B., McCauley,P., Boutell,J.M., Antonio,M.D. and Balasubramanian,S. (2019) Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res.*, **47**, 3862–3874.
53. Shen,J., Varshney,D., Simeone,A., Zhang,X., Adhikari,S., Tannahill,D. and Balasubramanian,S. (2021) Promoter G-quadruplex folding precedes transcription and is controlled by chromatin. *Genome Biol.*, **22**, 143.
54. Schaffitzel,C., Berger,I., Postberg,J., Hanes,J., Lipps,H.J. and Plückthun,A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Styloynchia lemnae macronuclei*. *Proc. Natl. Acad. Sci. U. S. A.*, **98**, 8572–8577.
55. Galli,S., Melidis,L., Flynn,S.M., Varshney,D., Simeone,A., Spiegel,J., Madden,S.K., Tannahill,D. and Balasubramanian,S. (2022) DNA G-Quadruplex Recognition In Vitro and in Live Cells by a Structure-Specific Nanobody. *J. Am. Chem. Soc.*, **144**, 23096–23103.
56. Zheng,K., Zhang,J., He,Y., Gong,J., Wen,C., Chen,J., Hao,Y., Zhao,Y. and Tan,Z. (2020) Detection of genomic G-quadruplexes in living cells using a small artificial protein. *Nucleic Acids Res.*, **48**, 11706–11720.
57. Yu,Z., Spiegel,J., Melidis,L., Hui,W.W.I., Zhang,X., Radzevičius,A. and Balasubramanian,S. (2023) Chem-map profiles drug binding to chromatin in cells. *Nat. Biotechnol.*, **41**, 1265–1271.

58. Hänsel-Hertsch,R., Spiegel,J., Marsico,G., Tannahill,D. and Balasubramanian,S. (2018) Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nat. Protoc.*, **13**, 551–564.
59. Henikoff,S., Henikoff,J.G., Kaya-Okur,H.S. and Ahmad,K. (2020) Efficient chromatin accessibility mapping in situ by nucleosome-tethered tagmentation. *eLife*, **9**, e63274.
60. Pavlova,I., Barinov,N., Novikov,R., Severov,V., Iudin,M., Vedekhina,T., Larin,A., Babenko,V., Aralov,A., Gnuchikh,E., *et al.* (2023) Modeling G4s in chromatin context confirms partial nucleosome exclusion and reveals nucleosome-disrupting effects of the least selective G4 ligands. *Biochimie*, **204**, 8–21.
61. Esnault,C., Garcia-Oliver,E., Aabidine,A.Z.E., Robert,M.-C., Magat,T., Gawron,K., Basyuk,E., Karpinska,M., Pigeot,A., Cucchiari,A., *et al.* (2023) G-quadruplexes are promoter elements controlling nucleosome exclusion and RNA polymerase II pausing. bioRxiv doi: <https://doi.org/10.1101/2023.02.24.529838>, 24 February 2023, pre-print: not peer-reviewed.
62. Cree,S.L., Chua,E.W., Crowther,J., Dobson,R.C.J. and Kennedy,M.A. (2020) G-quadruplex structures bind to EZ-Tn5 transposase. *Biochimie*, **177**, 190–197.
63. Stoltenburg,R., Krafčiková,P., Víglaský,V. and Strehlitz,B. (2016) G-quadruplex aptamer targeting Protein A and its capability to detect *Staphylococcus aureus* demonstrated by ELONA. *Sci. Rep.*, **6**, 33812.
64. Vorlíčková,M., Kejnovská,I., Sagi,J., Renčiuk,D., Bednářová,K., Motlová,J. and Kypr,J. (2012) Circular dichroism and guanine quadruplexes. *Methods*, **57**, 64–75.

FIGURE LEGENDS

Figure 1. Untargeted CUT&Tag has a consistent genome-wide enrichment. **(A)** Histogram of the number of peaks called by either MACS2 or SEACR that are shared by the specified number of datasets. **(B)** Percentage of unique or shared peaks between peaks of enrichment in untargeted ChIP-seq (48), untargeted CUT&RUN (30), and untargeted CUT&Tag samples. **(C)** Read counts for untargeted CUT&Tag datasets at a representative locus on chromosome 12 overlaid onto the untargeted CUT&Tag consensus peaks and ATAC-seq signal (average of 3 biological replicates) from K562 cells.

Figure 2. Untargeted CUT&Tag peaks colocalize with regulatory regions and accessible chromatin. **(A)** Proportions of annotations overlapping with untargeted CUT&Tag peaks using ChIPseeker (45). **(B)** Proportions of annotations overlapping with untargeted CUT&Tag peaks using HOMER (46). Metaplots of ATAC-seq signal (ENCSR868FGK; average of 3 biological replicates) at peaks of signal enrichment in untargeted CUT&Tag **(C)**, CUT&RUN (30) **(D)**, and ChIP-seq (48) **(E)** samples.

Figure 3. Untargeted CUT&Tag peaks colocalize with mapped G4s. **(A)** Top 5 known motifs and associated transcription factors (TF) present in untargeted CUT&Tag peaks identified by HOMER (46). All p-values $< 1 \times 10^{-125}$ by binomial test. **(B)** Top 5 *de novo* motifs and associated transcription factors (TF) present in untargeted CUT&Tag peaks identified by HOMER (46). All p-values $< 1 \times 10^{-20}$ by binomial test. **(C)** Proportion of untargeted CUT&Tag peaks present in the specified number of datasets that overlap a G4 mapped by BG4 CUT&Tag (27), BG4 ChIP-seq (53), G4Access (22), or mapped *in vitro* by G4-seq (52) or *in silico* by pqsfinder (51). **(D)** Proportion of overlap between untargeted CUT&Tag peaks and BG4 CUT&Tag (27) and/or BG4 ChIP-seq (53).

Figure 4. Enrichment of BG4 CUT&Tag over untargeted CUT&Tag does not increase at G4s in K562 cells. **(A-D)** BG4 or untargeted CUT&Tag signal (27) read counts were quantified at BG4 CUT&Tag peaks that overlap or do not overlap a G4 mapped by BG4 ChIP-seq (53) **(A)**, G4Access (22) **(B)**, G4-seq (52) **(C)**, or pqsfinder (51) **(D)** in K562 cells. Data are the average of 3 biological replicates, and the median and interquartile ranges are plotted in red. **(E)** ATAC-seq read counts at BG4 CUT&Tag peaks that overlap or do not overlap a G4 mapped by the

indicated method are plotted. Data are the average of 3 biological replicates, and the median and interquartile ranges are plotted in black. **(F)** Ratio of read counts of BG4 CUT&Tag (27) and untargeted CUT&Tag (27) quantified at individual CUT&Tag peaks that overlap or do not overlap a G4 mapped by the indicated method. Data are the average of 3 biological replicates, and the median and interquartile ranges are plotted in red. **(G)** Read counts (signal traces) and called peaks (rectangles) of ATAC-seq, BG4 CUT&Tag (27), untargeted CUT&Tag (27), G4Access (22), G4Access input (22), BG4 ChIP-seq (53) and ChIP-seq input (53) from K562 cells with G4-seq (52) and pqsfinder-predicted G4s (51) at a representative locus on chromosome 1. ATAC-seq, BG4 CUT&Tag, and untargeted CUT&Tag reads are the average of 3 biological replicates, G4Access reads are the average of 2 biological replicates, and G4Access input, BG4 ChIP-seq, and ChIP-seq input are data from 1 biological replicate. BG4 ChIP-seq peaks are the average of two biological replicates.

Figure 5. Enrichment of BG4 CUT&Tag over untargeted CUT&Tag does not increase at G4s within accessible chromatin in K562 cells. **(A)** Read counts of BG4 CUT&Tag (27) and untargeted CUT&Tag (27) from K562 cells at ATAC-seq peaks that overlap or do not overlap a mapped G4 (22, 51–53). Data are the average of 3 biological replicates, and the median and interquartile ranges are plotted in red. **(B)** Ratio of read counts of BG4 CUT&Tag (27) and untargeted CUT&Tag (27) from K562 cells quantified at ATAC-seq peaks that overlap or do not overlap a G4 mapped by the indicated method (22, 51–53). Data are the average of 3 biological replicates, and the median and interquartile ranges are plotted in red. **(C)** Read counts (signal traces) and called peaks (rectangles) of H3K4me3 CUT&Tag (24), H3K27me3 CUT&Tag (24), ATAC-seq, BG4 CUT&Tag (27), untargeted CUT&Tag (27), G4Access (22), G4Access input (22), BG4 ChIP-seq (53), and ChIP-seq input (53) from K562 cells with G4-seq (52) and pqsfinder-predicted G4s (51) at a representative locus on chromosome 4. Read counts of H3K4me3 CUT&Tag, H3K27me3 CUT&Tag, G4Access input, BG4 ChIP-seq, and ChIP-seq input are from 1 biological replicate. Read counts from G4Access are the average of 2 biological replicates. Read counts from ATAC-seq, BG4 CUT&Tag, and untargeted CUT&Tag are the average of 3 biological replicates. BG4 ChIP-seq peaks are the average of two biological replicates.