UC San Diego
UC San Diego Previously Published Works

Title

G-Quadruplex DNA and Other Non-Canonical B-Form DNA Motifs Influence Productive and Latent HIV-1 Integration and Reactivation Potential

Permalink

https://escholarship.org/uc/item/1j8098jt

Journal

Viruses, 14(11)

ISSN

1999-4915

Authors

Ajoge, Hannah O
Kohio, Hinissan P
Paparisto, Ermela
et al.

Publication Date

2022

DOI

10.3390/v14112494

Peer reviewed

Article

G-Quadruplex DNA and Other Non-Canonical B-Form DNA Motifs Influence Productive and Latent HIV-1 Integration and Reactivation Potential

Hannah O. Ajoge [1,†], **Hinissan P. Kohio** [1,†], **Ermela Paparisto** [1], **Macon D. Coleman** [1], **Kemen Wong** [1], **Sean K. Tom** [1], **Katie L. Bain** [1], **Charles C. Berry** [2], **Eric J. Arts** [1] **and Stephen D. Barr** [1,*]

[1] Schulich School of Medicine and Dentistry, Department of Microbiology and Immunology, Western University, Dental Sciences Building Room 3007, London, ON N6A 5C1, Canada

[2] Department of Family Medicine and Public Health, University of California San Diego, La Jolla, CA 92093, USA

[*] Correspondence: stephen.barr@uwo.ca

[†] These authors contributed equally to this work.

Abstract: The integration of the HIV-1 genome into the host genome is an essential step in the life cycle of the virus and it plays a critical role in the expression, long-term persistence, and reactivation of HIV expression. To better understand the local genomic environment surrounding HIV-1 proviruses, we assessed the influence of non-canonical B-form DNA (non-B DNA) on the HIV-1 integration site selection. We showed that productively and latently infected cells exhibit different integration site biases towards non-B DNA motifs. We identified a correlation between the integration sites of the latent proviruses and non-B DNA features known to potently influence gene expression (e.g., cruciform, guanine-quadruplex (G4), triplex, and Z-DNA). The reactivation potential of latent proviruses with latency reversal agents also correlated with their proximity to specific non-B DNA motifs. The perturbation of G4 structures in vitro using G4 structure-destabilizing or -stabilizing ligands resulted in a significant reduction in integration within 100 base pairs of G4 motifs. The stabilization of G4 structures increased the integration within 300–500 base pairs from G4 motifs, increased integration near transcription start sites, and increased the proportion of latently infected cells. Moreover, we showed that host lens epithelium-derived growth factor (LEDGF)/p75 and cleavage and polyadenylation specificity factor 6 (CPSF6) influenced the distribution of integration sites near several non-B DNA motifs, especially G4 DNA. Our findings identify non-B DNA motifs as important factors that influence productive and latent HIV-1 integration and the reactivation potential of latent proviruses.

Keywords: HIV; integration; latency; provirus reactivation; non-B DNA; guanine-quadruplex (G4) DNA; LEDGF/p75; CPSF6; reservoir

1. Introduction

An essential step in the life cycle of HIV-1 is the integration of its viral genome into the human genome. This event is permanent and leads to the life-long persistence of the virus within its host. Combination antiretroviral therapy (cART) suppresses productive HIV-1 replication in the infected individuals, thereby reducing the circulating virus to undetectable levels (reviewed in [1]). Despite this, resting memory CD4[+] T-cells harbor an integrated virus that persists in a transcriptionally silent state, referred to as latency [2,3]. HIV-1 remains latent indefinitely until it is reactivated by means that are not fully understood, but include the cessation of antiretroviral therapy, the development of antiretroviral resistance, or clinically directed 'shock and kill' therapy [4–6]. Latency presents a major obstacle in curing an individual of HIV-1 infection. This is in part due to the slow decay rate of an individual's latent reservoir after cART initiation, which has an estimated half-life of 44 months and an eradication timeline of >70 years in a patient [7,8]. Although

the size of the latent pool is under much debate, modeling studies have suggested that the expansion and contraction of latently infected cells can generate low-level persistent viremia and intermittent viral blips that can replenish the latent reservoir [9–11]. As such, the characterization of the latent reservoir is essential for the eradication of the virus from the body.

Multiple mechanisms have been attributed to establishing and maintaining proviruses in a latent state and are likely not mutually exclusive. For example, the site and orientation of the integration, the availability of the cellular transcription factors and viral proteins, the epigenetic regulation of the HIV-1 promoter, and the microRNA regulation of chromatin remodeling and targeting of mRNAs have been shown to contribute to latency (reviewed in [12,13]). 'Shock and kill' strategies have been proposed to flush out latent HIV-1 reservoirs to effect a cure. The main objective of these strategies is to facilitate the reactivation of HIV-1 expression from latent reservoirs, which are then destroyed through either natural means (e.g., immune response and viral cytopathogenicity) or artificial means (e.g., drugs and antibodies) [14]. Many latency reversing agents have been used for reactivation, including physiological stimuli, chemical compounds (phorbol esters), histone deacetylase (HDAC) inhibitors, p-TEFb activators, and antibodies (e.g., anti-CD3); however, these agents fail to reactivate the entire pool of the latently infected cells, highlighting the importance of a single integration site in an infected cell [15–18]. Although somewhat controversial, several reports suggest that this failure is due to genomic location-driven differences in HIV-1 expression [5,16,18–22]. Specific integration sites are also associated with the clonal expansion of latently infected cells [23]. Clonally expanded cells have been shown to produce infectious HIV-1 in vivo, however, this may not be the case in all infected individuals [24,25]. In fact, the majority of chromosomally integrated HIV-1 proviruses are defective in infected individuals [26]. A better understanding of the molecular mechanisms contributing to the establishment and maintenance of latency will aid current eradication strategies.

Much of the early retroviral integration site analyses focused on the most frequent integration events to better understand the genomic environment surrounding sites that result in a productive infection. The cellular proteins lens epithelium-derived growth factor (LEDGF)/p75 and cleavage and polyadenylation specificity factor 6 (CPSF6) are important factors in determining HIV-1 integration site selection, particularly into transcriptionally active genes. Notably, CPSF6 directs the preintegration complex away from heterochromatin at the periphery of the nucleus and towards the gene-dense chromosomal regions, whereas LEDGF/p75 primarily functions to position the preintegration complex along gene bodies [27–34]. Comparatively, there are fewer integration site analyses with respect to latent infection. This is likely attributed to the rarity of latent integration events and the small number of cells comprising the latent reservoir. With improved sequencing technology and methods to isolate latently infected cells, the genomic environment surrounding latent proviruses has been recently studied in more detail to identify the common genomic predictors of a latent integration. For example, using five different in vitro models of latency, Sherrill-Mix and colleagues identified some genomic features (e.g., histone acetylation and alphoid repeats) which were associated with proviral expression in individual models [21]. Although they did not identify a common genomic predictor of latency across all of the models, the authors proposed that the features affecting latency were highly local and heterogeneous. Indeed, other groups have also shown that the latent reservoir is heterogeneous in nature [18,20,35]. The HIV-1 LTR promoter activity was also shown to be sensitive to the local chromatin environment and not controlled by DNA methylation or histone acetylation in cell lines [19,36]. Proviral insertion sites also affect the response to latency reversal agents (LRAs), where different LRAs can activate different subsets of proviruses in the latent population [18,20]. These different subsets are distinguishable in terms of the chromatin functional states and only represented <5% of cells carrying a latent provirus [18]. Recently, it was shown that individuals on prolonged antiretroviral therapy undergo selection of the intact proviruses with features of deep viral latency [37].

We previously identified non-B DNA motifs as a factor that potentially influences the HIV-1 integration site targeting [38]. Non-B DNA motifs are abundant in the human genome and form secondary structures using non-canonical Watson–Crick base pairing [39]. At least 10 non-B DNA conformations exist including A-phased repeats, inverted repeats, direct repeats, mirror repeats, short-tandem repeats, triplex repeats, G4, cruciform, slipped, and Z-DNA [39,40]. Several non-B DNA features preferentially act as the recipient of genetic information, stimulating homologous recombination >20-fold in human cells [41]. Some non-B DNA structures (e.g., G4, cruciform, triplex, and Z-DNA) have also been shown to potently silence the expression of adjacent genes [42–52]. In this study, we analyzed the HIV-1 integration site profiles of productively and latently infected cells and identified a correlation between non-B DNA motifs, provirus integration sites, and the reactivation potential of latent proviruses.

2. Materials and Methods

2.1. Cell Propagation, Virus Production and Infection

HEK293T cells (female) were obtained from the American Type Culture Collection and were maintained in standard Dulbecco's Modified Eagle Medium (DMEM) or phenol red-free DMEM at 37 °C with 5% CO_2 in a humidified incubator. All the media were supplemented with 10% heat-inactivated fetal bovine serum (FBS), 100 U/mL of penicillin, and 100 μg/mL of streptomycin. Pseudotyped HIV/VSV-G was generated by co-transfecting HEK293T cells with the plasmids p156RRLsinPPTCMVGFPWPRE (encoding the HIV vector segment), pCMVdeltaR9 (the packaging construct), and pMD.G (encoding the VSV-G envelope), as previously described [53]. The 293T cells, which were readily infected by HIV/VSV-G, were infected for 5 h with pseudotyped HIV-1/VSV-G in the presence of 10 μg/mL of polybrene (Sigma-Aldrich, St. Louis, MO, USA, #H9268-5G). The plasmids pHIV$_{GKO}$ and pHIV$_{GKO-DU3LTR}$ were kindly provided by Dr. Eric Verdin (Buck Institute). The pseudotyped HIV$_{GKO}$ and HIV$_{GKO-DU3LTR}$ viruses were similarly generated by co-transfecting HEK293T cells with pHIV$_{GKO}$ or pHIV$_{GKO-DU3LTR}$ and pMD.G for 48 h. Jurkat cells, which were readily infected by HIV$_{GKO}$ and amenable to growth in suspension for 96 h, were infected via spinoculation at 3000 rpm at room temperature for 3 h in the presence of 10 μg/mL of polybrene (Sigma-Aldrich, #H9268-5G). Following the spinoculation, the virus was removed, and fresh medium was added to the well for 4 days.

2.2. Drug Treatment and Infection for Integration Site Analysis

BRACO-19 hydrochloride (BRACO-19) was purchased from Sigma-Aldrich (#SML0560-5MG). TMPyP4 was purchased from Calbiochem-EMD Millipore (#613560-25MG). 0.5×10^6 HEK293T cells were plated in 6-well plates and left untreated (control) or treated for 24 h with either BRACO-19 (0, 1, 3, and 32 μM) or TMPyP4 (0, 0.5, 1, and 8 μM). The BRACO-19 and TMPyP4 concentrations were established from previously used concentrations [54–56]. The cells were then infected for 5 h with pseudotyped HIV-1/VSV-G in the presence of 10 μg/mL of polybrene (Sigma-Aldrich, #H9268-5G). The medium was changed 5 h post-infection and the cells were treated anew with either BRACO-19 or TMPyP4 for 24 h. The genomic DNA was extracted from the cells as per the manufacturer's instructions using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany, #69504) and processed for the integration site profile analyses.

2.3. HIV Integration Site Library and Computational Analysis

Integration sites were determined from the sequence junction of the LTR and human genome sequences. The genomic DNA was processed for integration site analysis and sequenced using the Illumina MiSeq (San Diego, CA, USA) platform, as previously described [38,57]. Fastq sequencing reads were quality trimmed and the unique integration sites were identified using our in-house bioinformatics pipeline, which was described previously and is now called the Barr Lab Integration Site Identification Pipeline (BLISIP version 2.9) [38]. BLISIP version 2.9 includes the following updates: bedtools (v2.25.0),

bioawk (awk version 20110810), bowtie2 (version 2.3.4.1), and restrSiteUtils (v1.2.9). The HIV-1 LTR-containing fastq sequences were identified and filtered by allowing up to a maximum of five mismatches with the reference NL4-3 LTR sequence. The LTR sequences matching any region of the human genome (GRCh37/hg19) were discarded. Sites located in various common genomic features and non-B DNA motifs were quantified. The heatmaps were generated using our in-house python program BLISIP Heatmap (BLISIPHA v1.0), which calculates the fold enrichment of sites in the desired distance bin from each feature compared to that of the desired control dataset. The matched random control datasets were generated from the hg19 reference genome to account for restriction site bias in the preparation of the libraries that used restriction enzymes in their construction, as previously described [38,58]. The random control datasets were generated from the hg19 reference genome using the random tool of BEDTools v2.28 for the libraries generated by random fragmentation (e.g., shearing) [59]. The sites that could not be unambiguously mapped to a single region in the genome were excluded from the study. All non-B DNA motifs were defined according to the previously established criteria [60]. The G-quadruplex algorithm identifies four or more individual G-runs of at least three nucleotides in length. The algorithm requires at least one nucleotide between each run and considers up to seven nucleotides as the spacer, including the guanines [60].

2.4. MTT Assay

HEK293T or Jurkat cells were treated with increasing concentrations of BRACO-19 (0, 1, 3, and 32 μM) or TMPyP4 (0, 0.5, 1, and 8 μM) for 5 days at 37 °C with 5% CO_2. The cell metabolic activity was measured using the MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide) kit (Thermo Fisher Scientific, Waltham, MA USA #M6494) according to the manufacturer's instructions. The absorbance of the plates was read at 540 nm using the Agilent Biotek Epoch microplate spectrophotometer (Agilent Technologies, Santa Clara, CA, USA) plate reader and the Gen5 version 2.06 analysis software. The percent viability was calculated relative to the untreated control cells.

2.5. Confocal Immunofluorescence Microscopy

Jurkat cells were treated with 0, 8, or 32 μM of BRACO-19 for 24 h. Intracellular G4 structure antibody staining protocols were adapted from previously published protocols [61,62]. The cells were then fixed with 4% paraformaldehyde (Electron Microscopy Science, Hatfield, PA, USA) for 10 min and permeabilized with 0.5% Tween 20 in 1× PBS. The blocking was performed overnight at 4 °C with 5% goat serum (Wisent Bio Products, St-Bruno, Quebec, Canada) in 1× PBS. The cells were incubated for 24 h with an anti-G4(1H6) antibody (Millipore Sigma-Aldrich, Oakville, Ontario, Canada) at a ratio of 1:100. The cells were washed 5 times with 0.1% Tween 20 in 1× PBS (PBST). Alexa Fluor 594 (Thermo Fisher Scientific Life Technologies, Wyman Street, Waltham, MA USA) secondary antibody staining was performed at a ratio of 1:1000 for 2 h at room temperature. The cells were washed 3 times with PBST and 2 times with 1× PBS. The cell nuclei were stained with a Hoechst stain for 2 min at room temperature. The cells were washed 5 times in 1× PBS and maintained in 1× PBS for the duration of the imaging. As a control, the Jurkat cells were incubated for 24 h with 32 μM of BRACO-19, fixed, permeabilized, blocked, and stained with a secondary antibody as described above (in the absence of the primary anti-G4(1H6) antibody). The imaging was performed with the Leica TCS SP8 confocal microscope using the LAS X Life Science software (Leica Microsystems, Concord, Ontario, Canada). To quantify the G4 structures in the nucleus, a mask of each Hoechst-stained nuclei was generated and superimposed onto the fluorescent G4 channel of the same image using Fiji (ImageJ version 2.1.0/1.53c) [63]. The G4 fluorescence was measured for all of the nuclei in the field of view with an integrated intensity above 800. At least 430 cells from 40 fields of view per condition were measured from two independent experiments.

2.6. Flow Cytometry

Four days after infection with the HIV$_{GKO}$ constructs, the Jurkat cells were stained with Zombie NIR^{TM} (BioLegend, #423105) to label the live cells and have them fixed with paraformaldehyde to a final concentration of 2%. The data were collected with a FACSCanto (Becton Dickinson, Franklin Lakes, NJ, USA) and the analyses were performed with FlowJo version 10.0.7 software (Flowjo, Ashland, OR, USA). The proportions of latent or productive integrations were calculated by dividing the percentage of the (GFP−,mKO2+) or (GFP +,mKO2+) cells (respectively) by the total percentage of positive cells ((GFP−,mKO2+)+(GFP+,mKO2+)). The percentage of (GFP−,mKO2−) or (GFP+,mKO2−) cells was not included when determining the proportion of latent and productive cells.

2.7. Datasets

Integration site datasets used in this study were independently analyzed using our in-house bioinformatics pipeline BLISIP v2.9. The Battivelli dataset was obtained from the "Integration Sites—Source Data" from reference [18]. The Achuthan dataset was obtained from the NCBI SRA using the accession number SRP132583, as described in [34]. All of the genomic sites in each dataset that hosted two or more sites (i.e., identical sites) were collapsed into one unique site for the analysis herein.

2.8. Statistical Analyses

All of the statistical tests were performed as described in the figure legends using Graphpad Prism 6 (Graphpad Software, San Diego, CA, USA).

3. Results

3.1. Reactivation Potential of HIV-1 Proviruses Correlates with Integration Site Placement near Non-B DNA Motifs

Recently, it was shown by Battivelli et al. (2018) that HIV-1 integration sites were distinguishable with respect to chromatin functional states and that these locations correlated with latency reactivation [18]. Given that several non-B DNA structures can influence chromatin organization and gene expression (e.g., G4, cruciform, triplex, and Z-DNA), we asked if latent provirus reactivation potential correlated with integration site placement near non-B DNA motifs. We defined 'near' as a window of the +/− 500 base pairs (bp) from the non-B DNA motif. The reason for selecting this distance was three-fold. First, we selected a distance with potential functional significance. For example, the region of the genome proximal to the TSS is essential for transcriptional regulation, and G4 motifs have been reported to be strongly enriched in the transcriptional regulatory region (defined as 500 bp upstream and downstream of the TSS) in warm-blooded animals [64,65]. Second, non-B DNA structures have been shown to influence positioning of the flanking nucleosomes encompassing 500 bp upstream and downstream of the structure, which equates to the length of DNA that wraps around approximately three nucleosomes on each side of the non-B DNA motif [66]. Third, a short distance was also selected to minimize the chance of multiple non-B DNA motifs being captured in the 500 bp window.

The Battivelli integration site dataset was previously generated from primary CD4+ T cells infected with a dual-fluorescence HIV-1 reporter virus (HIV$_{GKO}$) (Figure 1A) [18]. This virus was designed for the quantification and purification of a large number of latently infected cells by flow cytometry. Infection of the cells with HIV$_{GKO}$ produced three populations of infected cells (Figure 1B). The first population contained productively infected cells (PIC) that expressed a codon-switched green fluorescent protein (GFP) under the control of the HIV-1 promoter in the 5′ long terminal repeat (LTR) and a distinct unrelated fluorescent protein mKO2 under the control of the EF1α promoter (GFP+, mKO2+) (Figure 1A,B). The second population contained latently infected cells (LIC) that express only mKO2 (GFP−, mKO2+). This LIC population was further divided into two additional sub-populations based on their αCD3/CD28 reactivation potential, cells that could be reactivated (RLIC), or

cells that could not be reactivated (NRLIC) (Figure 1B). The third population of cells contained uninfected, dead, or defective cells (GFP+, mKO2−), and cells latent for both markers (GFP−, mKO2−). The third population of cells was excluded from further analyses.
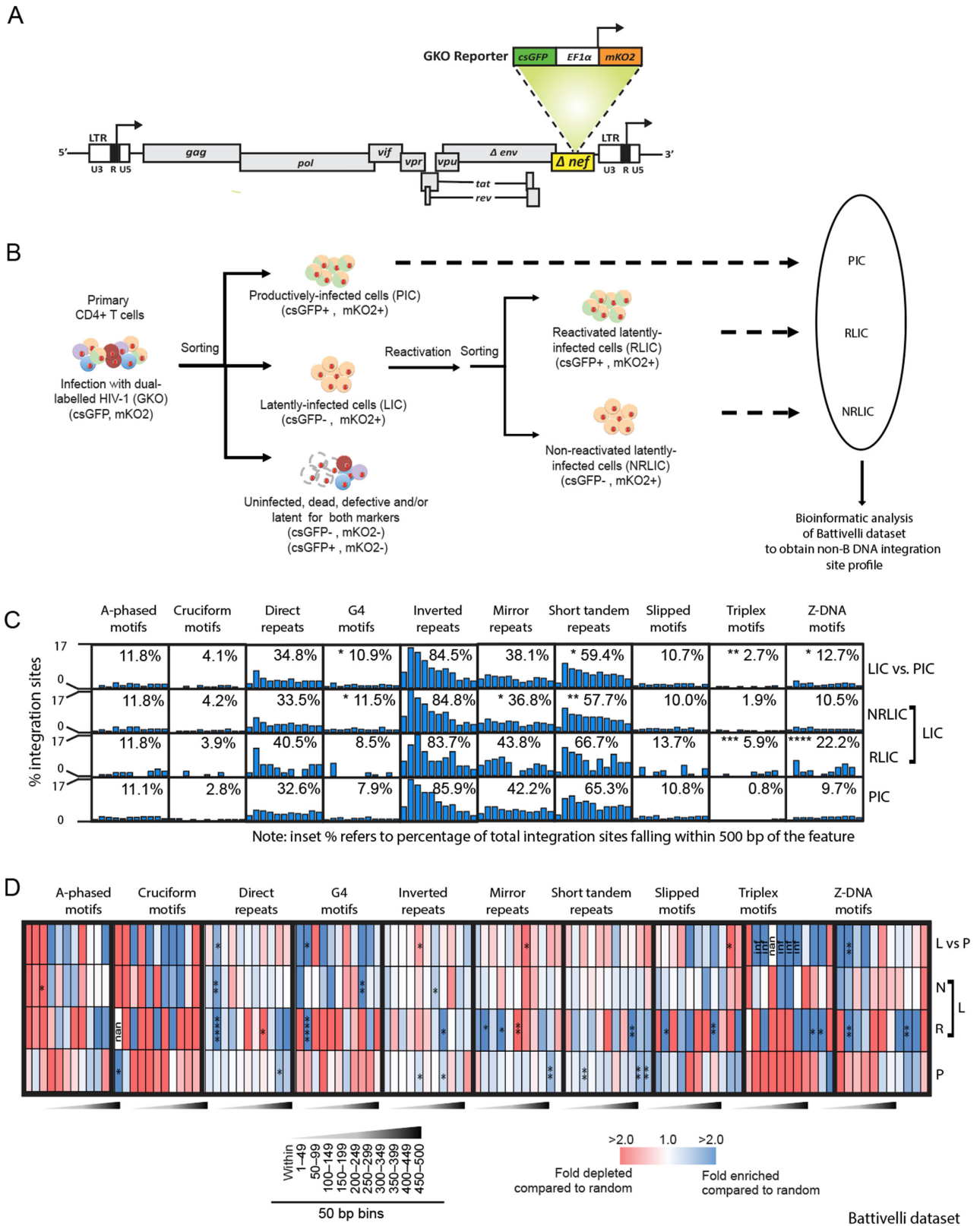

Figure 1. Integration near specific non-B DNA is associated with proviral reactivation in latently infected cells. (**A**), Schematic depicting the HIV-1 vector used to generate productively and latently

infected cells. (**B**), Strategy used in the Battivelli study to isolate productively infected cells (PIC), latently infected cells (LIC), reactivatable latently infected cells (RLIC), and non-reactivatable latently infected cells (NRLIC) cell populations. (**C**), Graphs show the percentage of total unique HIV-1 integration sites located in or within 500 bp of various non-B DNA motifs (distributed in 50 bp bins). Inset numbers show the percentage of total unique integration sites falling within 500 bp of the non-B DNA motif. Statistical analysis is with respect to PIC. (**D**), Heatmaps show the fold enrichment (blue) or depletion (red) of integration sites at each distance interval from the non-B DNA motif compared to the random control. The LIC population was compared to the PIC population. Fisher's exact test; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. Infinite number (inf), 1 or more integrations were observed when 0 integrations were expected by chance. Nan, not a number (0 integrations were observed and 0 were expected by chance). PIC (P), RLIC (R), NRLIC (N), LIC (L), LIC versus PIC (L vs P).

We generated integration site profiles from the previously published Battivelli integration site dataset to determine if the reactivation potential of the HIV-1 proviruses correlated with integration site placement near non-B DNA motifs. For comparison, our integration site profile analysis was in agreement with the conclusions from the Battivelli study showing that the integration sites for the LIC population (NRLIC + RLIC) were enriched in regions of heterochromatin (e.g., satellite DNA and lamin-associated domains (LADs)) at a higher degree than the PIC population (Figure S1 and Tables S1 and S2). Analysis of our newly generated non-B DNA integration site profile showed that the majority of productive integration sites (PIC) were enriched near direct, inverted, mirror, and short tandem repeats (ranging from 33–86% of the total integration sites), followed by A-phased, slipped, G4, and Z-DNA motifs (8–11%), and a smaller percentage of sites within cruciform and triplex motifs (1–3%) (Figure 1C and Table S2). The LIC population exhibited a similar profile to that of the PIC population with the notable exception of reduced integration near the short tandem repeats (59% compared to 65%) and increased integration near the G4, triplex, and Z-DNA motifs (3–11% compared to 1–3%) (Figure 1C and Table S2).

Analysis of the RLIC and NRLIC populations separately showed that sites in the RLIC population were more enriched near triplex (6%) and Z-DNA (22%) motifs compared to the NRLIC population (2%, $p < 0.05$; and 11%, $p < 0.001$; respectively) (Figure 1C and Table S2). In addition, sites in the NRLIC population were enriched near the G4 motifs (12%) and reduced near mirror repeats (37%) and short tandem repeats (58%), compared to the RLIC population (9%, $p > 0.05$; 44%, $p > 0.05$; and 67%, $p < 0.05$; respectively).

To assess the distribution of the integration sites within the 500 bp window around the non-B DNA motifs, we sub-divided the 500 bp distance into 50 bp bins and quantified the integration sites in each bin (Figure 1C,D and Table S2). The distribution of sites in the PIC population were mostly uniformly distributed among the bins with an enrichment of sites in the bins that were more distal to direct, inverted, mirror, and short tandem repeats (~300–500 bp). In contrast, sites in the LIC population were enriched in bins more proximal (~1–100 bp) to direct repeats and G4, triplex, and Z-DNA motifs (Figure 1C,D and Table S2). Comparison of the RLIC and NRLIC populations separately showed that the RLIC population exhibited increased integration in the bins 1–50 bp away from direct repeats, mirror repeats, and G4, slipped, and Z-DNA motifs, and increased integration in the bins 350–500 bp away from the inverted repeats, short tandem repeats, triplex, and Z-DNA. Notably, the NRLIC population exhibited increased integration in the bin 350–399 bp away from G4 motifs compared to the RLIC population.

Together, these data show that different HIV-1 integration site biases exist towards non-B DNA motifs between productively and latently infected cells. Moreover, the reactivation of latent virus correlated with increased integration near mirror repeats, short tandem repeats, triplex, and Z-DNA motifs, and inversely correlated with integration near G4 motifs.

3.2. Pretreatment of Cells with G4 Ligands Alters HIV-1 Integration Targeting of G4 DNA

Given that the reagents to study G4 structures are more readily available than those for cruciform, triplex, and Z-DNA, we chose to further examine the influence of G4 DNA on HIV-1 integration site selection in vitro. G4 structures are four-stranded secondary structures containing four guanine bases that form a G-tetrad in a planar arrangement (Figure 2A) [40,67–69]. To determine the influence of G4 structures on integration site selection in the human genome during acute HIV-1 infection, we performed an experiment utilizing G4 structure-destabilizing and -stabilizing ligands TmPyP4 and BRACO-19, respectively (Figure 2B,C). The cationic porphyrin 5,10,15,20-tetra(N-methyl-4-pyridyl)porphin (TMPyP4) interacts with and destabilizes long-loop non-telomeric G4 structures while, paradoxically, stabilizing short-loop G4 structures located in the telomeric DNA [55,56,70–73]. We asked if the destabilization of non-telomeric G4 structures reduced integration near G4 DNA motifs. The control 293T cells or the cells treated with increasing concentrations of TMPyP4 were infected with HIV-1 pseudotyped with the vesicular stomatitis virus G envelope glycoprotein (HIV/VSV-G).

No significant reduction in the cell viability was detected (Figure 2D). Integration was significantly reduced within 500 bp of the G4 DNA motifs with increasing concentrations of TMPyP4 (Figure 2E and Tables S1 and S3). This reduced integration was observed despite an unexpected increase in the integration at low concentrations of TMPyP4 in a region 250–500 bp away from G4 DNA motifs.

BRACO-19 is a 3,6,9-trisubstituted acridine derivative that interacts with and stabilizes the G4 structures [54,72,74–77]. Although it was previously shown that BRACO-19 reduces but does not abolish reverse transcription at the template level due to the presence of a G4 motif at the 3′ end of the viral RNA genome [54], we asked if BRACO-19 treatment could alter integration targeting of the pre-integration complexes that are produced. Control 293T cells or cells treated with increasing concentrations of BRACO-19 were infected with HIV/VSV-G. No significant reduction in cell viability was detected after treatment with BRACO-19 (Figure 2F). Increasing concentrations of BRACO-19 resulted in a heterogeneous pattern of integration near the G4 motif where integration was significantly reduced within 100 bp and was significantly enriched 300–500 bp from the G4 DNA. Unexpectedly, integration was increased 150–199 bp from G4 DNA at low concentrations of BRACO-19, which then decreased at higher concentrations.

As a measure of specificity, integration in other commonly referenced genomic features showed that BRACO-19 treatment resulted in a 2.8-fold increase in integration sites near transcription start sites compared to the no drug control and TmPyP4-treated cells (Figure 2H). This finding was expected since TSS are rich in G4 DNA motifs [78–82]. No substantial change in the proportion of integrations sites within genes were observed with either treatment (Figure 2I). To determine if BRACO-19 or TmPyP4 treatment influenced the known preference of HIV-1 for integration into highly expressed genes, we analyzed a previously published gene expression profiling dataset for 293T cells (GSE2451). The genes were categorized into four different expression bins ranging from a low- to high-level of gene expression. Integration sites from the cells exhibited a significant enrichment in highly expressed genes regardless of the treatment. The analysis of the integration sites along each chromosome in the presence or absence of either treatment revealed no integration sites located in the telomeric regions, and thus did not appear to alter the normally low level of HIV-1 integration in telomeres (Figure S2 and Table S5). However, since the integration sites that could not be unambiguously mapped to a single region in the genome were excluded from analysis, it is possible that the sites in the telomeric repeats were present but undetected.

Together, these data show that pre-treatment of cells with G4 structure-stabilizing or -destabilizing ligands reduced integration within 100 bp of the motifs, whereas the stabilizing ligand BRACO-19 increased integration 300–500 bp away from the motif.
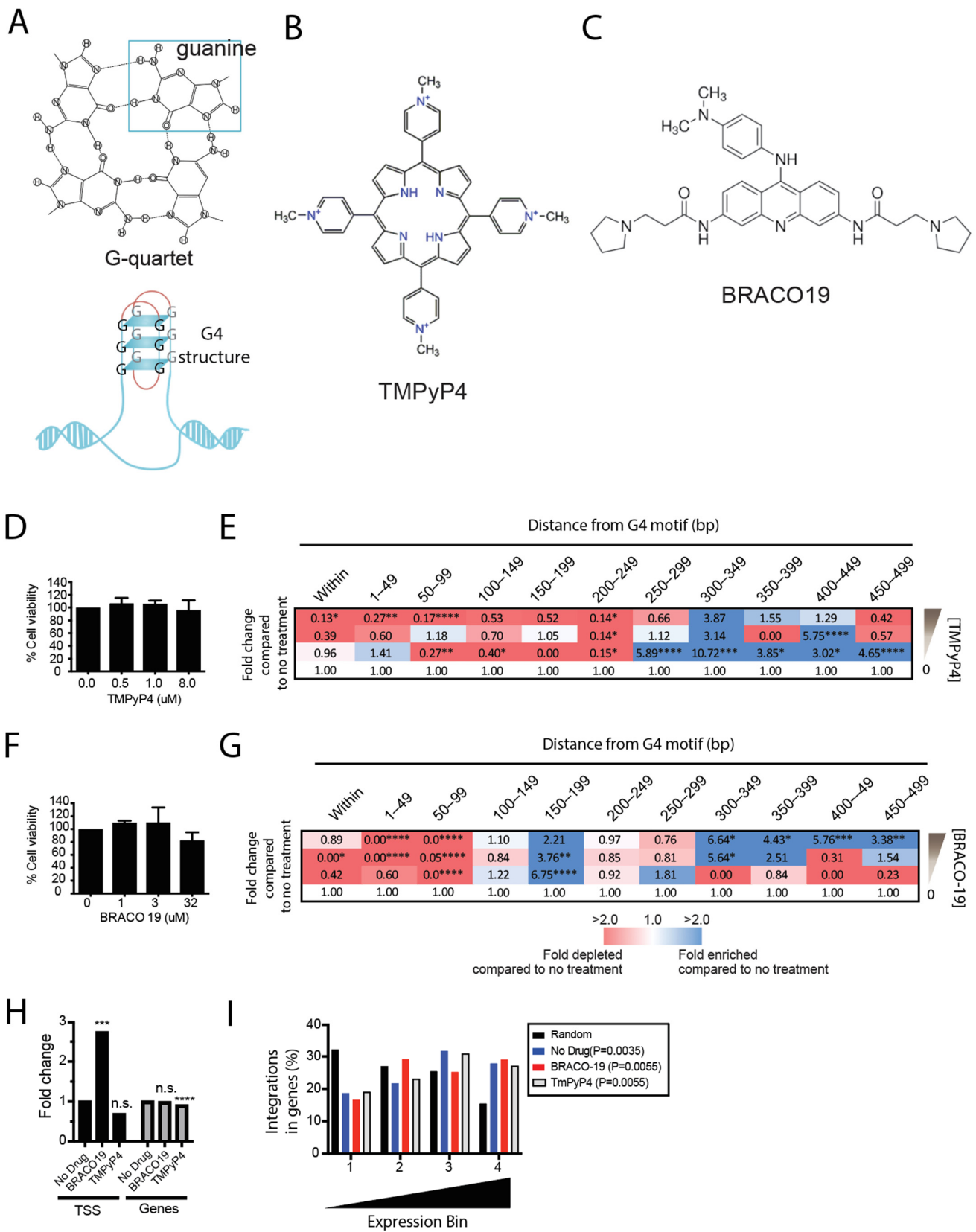
Figure 2. G4 structures influence integration site targeting in the genome. (**A**), Depiction of the nucleoside arrangement of a guanine-quartet and the G4 structure. (**B**), Structure of the TMPyP4 compound. (**C**), Structure of the BRACO-19 compound. (**D**), 293T cells were treated with increasing concentrations of TMPyP4 for 48 h and percent cell viability was determined using the MTT assay.

Data are represented as mean ± SEM of at least 3 independent experiments. (**E**), Heatmap shows the fold enrichment (blue) or depletion (red) of integration sites at each 50 bp distance interval from the G4 motif compared to the untreated infected control cells. Statistical analysis is with respect to the untreated control. (**F**), 293T cells were treated with increasing concentrations of BRACO-19 for 48 h and percent cell viability determined using the MTT assay. Data are represented as mean ± SEM of at least 3 independent experiments. (**G**), Heatmap shows the fold enrichment (blue) or depletion (red) of integration sites at each 50 bp distance interval from the G4 motif compared to the untreated infected control cells. (**H**), Fold change in the percentage of integration sites located within 500 bp of transcription start sites (TSS) and directly in genes before and after treatment with BRACO-19 (32 μM) or TmPyP4 (8 μM). Statistical analysis is with respect to the untreated control. Fisher's exact test; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. n.s., not significant. (**I**), Analysis of expression levels of genes targeted for integration with or without BRACO-19 (32 μM) or TmPyP4 (8 μM). Gene expression levels in 293T cells were obtained from the published GEO dataset GSE2451. A total of 22,288 genes were assayed and distributed into four equal bins by relative expression levels. The bin with the lowest average expression is at the left and the highest expression is at the right. Genes used as integration targets with or without treatment were distributed into their corresponding bins based on their expression levels and summed.

3.3. HIV-1 Favors G4 Structures with Long-Loops for Integration

Putative G4 structures are identified using the motif $G_xN_{y1}G_xN_{y2}G_xN_{y3}G_x$ [83–86]. The motif consists of four guanine tracts with three intervening loops (Figure 3A). In this expression, x represents the number of guanine nucleotides, Ny1–Ny3 represent the three intervening loops and can be categorized as short-loop G4 structures (1–7 nucleotides), or long-loop G4 structures (≥7 nucleotides) based on the number of nucleotides (N) in the loop. The typical loop lengths in the human genome are between 1 and 7 nucleotides [87]. The loop-length has been shown to play an important role in the G4 structure's stability and protein-binding specificity [85,88]. We asked if the HIV-1 integration sites observed in the PIC, LIC, RLIC, and NRLIC populations from the Battivelli study were biased towards short- or long-loop G4 DNA motifs and if loop-length correlated with reactivation potential of latently infected cells. The G4 DNA sequences were extracted from the Battivelli dataset and the average loop-lengths were compared for each of the three loops. No difference in the loop-lengths was observed between the PIC, RLIC, or NRLIC populations (Figure 3B). Notably, loop two was twice as long as loops one and three within each population of cells. These data indicate that HIV-1 integration is biased towards long-loop G4 motifs and that the loop length does not correlate with productive or latent integration, or the reactivation potential of latently infected cells.
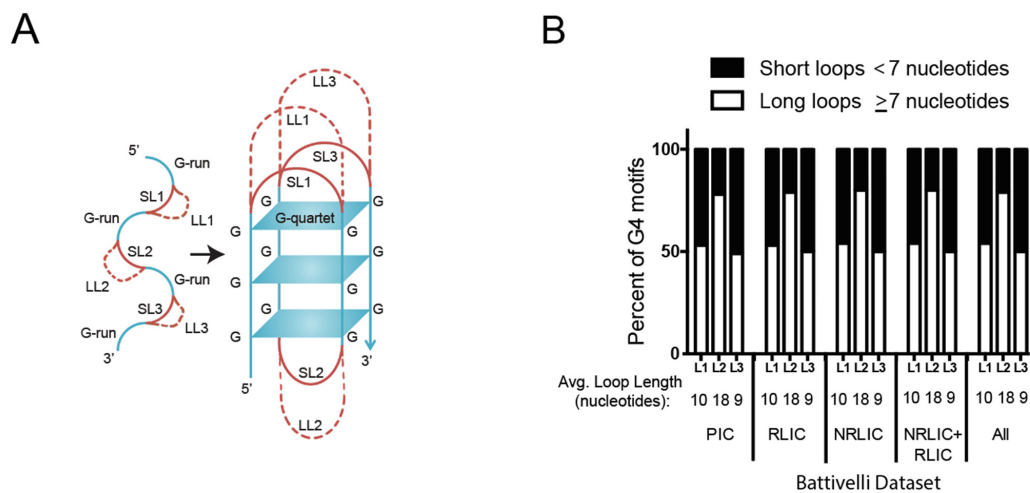

Figure 3. Integration in or near G4 motifs favors G4 structures with long-loops. (**A**), Schematic depicting a G4 motif (left) and G4 structure (right) consisting of four adjacent runs of two or more

guanines, with three loop regions of nucleotide subsequences (L1, L2, and L3) connecting the G-runs. Loops containing <7 nucleotides are considered short-loop (SL) G4s (solid red line), whereas ≥7 are considered long-loop (LL) G4s (dashed red line). (**B**), G4 motif sequences located within 500 bp of an integration site were isolated from the Battivelli datasets and the loop-lengths were calculated. The percentage of G4 motifs in each dataset classified as short-loops or long-loops were compared in the bar graphs. The average loop lengths for each of the three loops were calculated and are shown below each bar graph.

3.4. Stabilization of G4 Structures In Vitro Increases the Proportion of Latently Infected Cells

G4, cruciform, triplex, and Z-DNA have been shown to silence the expression of adjacent genes [42–52]. Given our finding that integration sites are enriched near G4 DNA motifs in latently infected cells and that altering the G4 DNA stability influences integration targeting of G4 DNA motifs, we asked if the stabilization of G4 structures in vitro increases the proportion of the latently infected cells in a pool of infected cells. Twenty-four hours prior to the infection, we treated the Jurkat cells with increasing concentrations of BRACO-19 to stabilize the intracellular G4 structures. The stabilization of nuclear G4 structures was verified by confocal immunofluorescence microscopy using anti-G4(1H6), which binds specifically to the G4 DNA structures [62,89]. As expected, nuclear G4 staining increased with increasing concentrations of BRACO-19 (Figure 4A,B). Some G4 DNA staining was also observed in the cytoplasm, which is consistent with previous observations and is likely attributed to the release of G4 DNA structures into the cytoplasm after DNA damage repair processes [90]. No auto-fluorescence was observed with BRACO-19 at the highest concentration used (32 μM) (Figure 4A).

To measure the number of productively and latently infected cells in a pool of infected cells, we infected Jurkat cells with the HIV_{GKO} dual-reporter virus (VSV-G pseudotyped), as described above (see Figure 1A,B) [18]. Productively infected cells were double-positive for GFP and mKO2 fluorescence (GFP+, mKO2+), whereas latently infected cells were positive for mKO2 fluorescence only (GFP−, mKO2+). Cells exhibiting only GFP expression (GFP+, mKO2−) were considered defective. Cells not expressing either marker (GFP−, mKO2−) comprise uninfected or dead cells, cells containing defective proviruses, and/or cells containing proviruses latent for both a GFP and mKO2 expression. As a control, we utilized the same HIV_{GKO} virus lacking the U3 promoter region of the 3′ LTR ($HIV_{GKO-DU3LTR}$), which results in an integrated virus devoid of the 5′ HIV U3 region (GFP−, mKO2+).

To test the effect of stabilizing G4 structures on the proportion of productively and latently infected cells, the Jurkat cells were treated with BRACO-19 for 24 h and then infected with HIV_{GKO} for 4 days. The live cells were identified by flow cytometry using Zombie NIR^{TM} staining and analyzed for GFP and mKO2 fluorescence. As expected, infection with the $HIV_{GKO-DU3LTR}$ control virus yielded >99% (GFP−, mKO2+) cells (Figure 4C). Additionally, as expected, increasing concentrations of BRACO-19 caused a reduction in, but not an abolishment of, infection due to its inhibitory activity on reverse transcription (Figure 4C,D). Analysis of the infected cells (mKO2+) showed that the proportion of productively infected cells (GFP+, mKO2+) decreased from 60% to 22%, whereas the proportion of latently infected cells (GFP−, mKO2+) increased from 40% to 78% ($p < 0.0001$, linear regression; $R^2 = 0.480$, DFn = 1, DFd = 44, F = 40.57,) (Figure 4D,E). In comparison, analysis of the infected cells treated with increasing concentrations of TmPyP4 showed a more modest effect compared to the BRACO-19 treatment, where the proportion of the productively infected cells (GFP+, mKO2+) decreased from 54% to 46%, and the proportion of the latently infected cells (GFP−, mKO2+) increased from 46% to 54% ($p < 0.0001$, linear regression; $R^2 = 0.264$, DFn = 1, DFd = 56, F = 20.06) (Figure 4F,G). This finding suggests that G4 DNA is not the only factor contributing to latent infection. Similar to BRACO-19, TmPyP4 also caused a reduction in, but not an abolishment of, infection (Figure 4F). The mechanism(s) underlying this anti-HIV activity of TmPyP4 has not been fully characterized. TmPyP4 may interfere with important interactions between a G4 structure located in the HIV DNA flap region and nucleocapsid protein during nucleocapsid assembly, and/or

it may interfere with the interstrand quadruplex formation in the dimerization of HIV RNA [91,92].

Together, these data show that the stabilization of nuclear G4 structures with BRACO-19 increases the number of latently infected cells and decreases the number of productively infected cells in a population of HIV-1 infected cells in vitro.
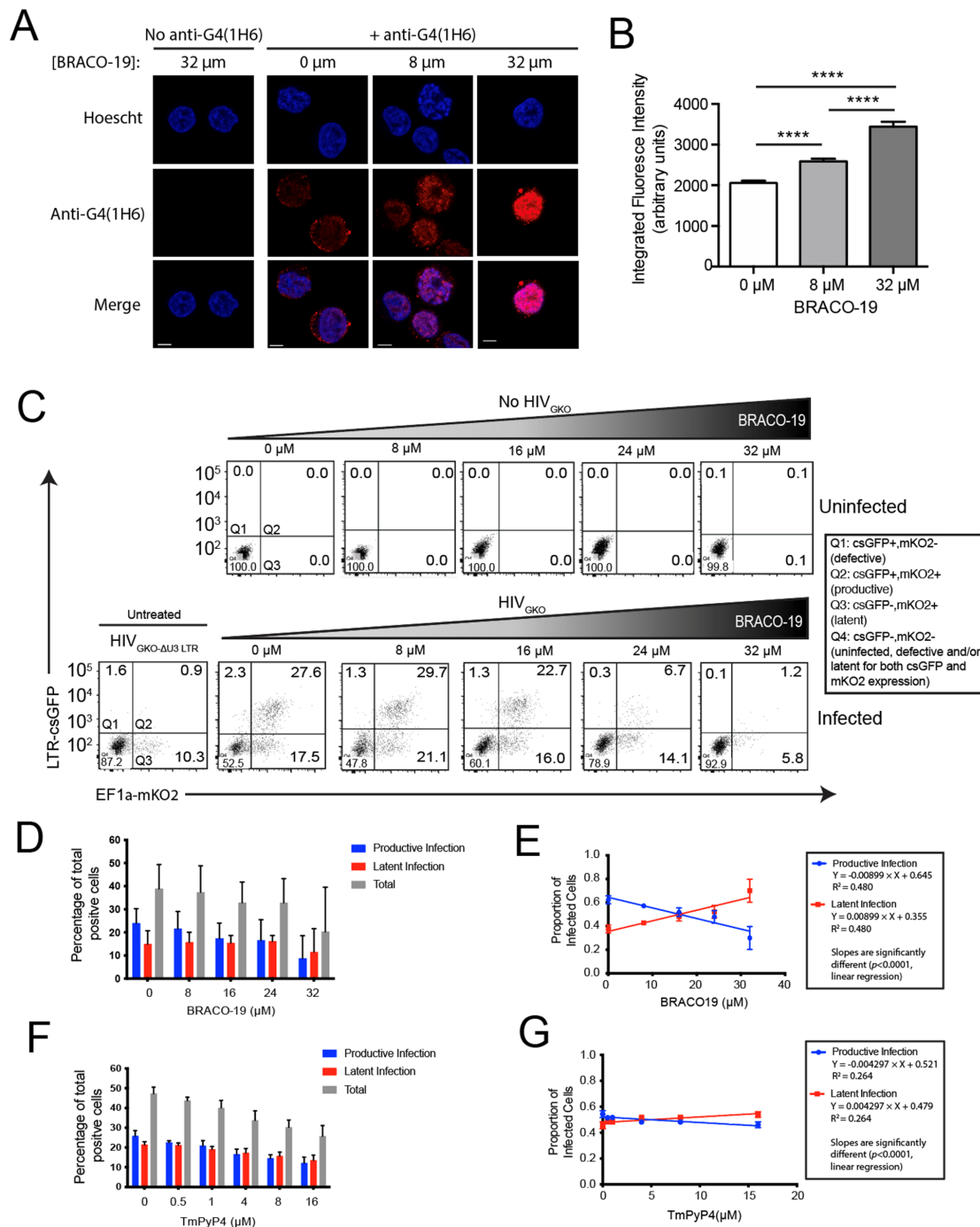

Figure 4. Stabilization of G4 structures increases the proportion of latently infected cells in vitro. (**A**), Jurkat cells were pre-treated for 24 h with increasing concentrations of BRACO-19 and then subjected to confocal immunofluorescence microscopy using anti-G4(1H6) antibody. Nuclei were identified with Hoechst staining. (**B**), Nuclear G4 structure content was quantified by creating a mask of Hoechst nuclear staining for all nuclei using FIJI (NIH). Each mask was overlaid onto the G4 staining channel of the same image and the G4 staining was measured for each nucleus in the field of

view with an integrated intensity above 800. Data shown represent the average (+/− S.E.M.) of at least 430 cells per condition from 40 fields of view from two independent experiments. One-way ANOVA with Tukey's multiple comparisons test; ****, $p < 0.0001$. (**C**), Jurkat cells were pre-treated for 24 h with increasing concentrations of BRACO-19 and infected with HIV_{GKO} or $HIV_{GKO\text{-}DU3LTR}$ for 96 h in the presence of drug. Cells were stimulated with anti-CD3/CD28. Live cells were gated on using Zombie NIR^{TM} staining and the percentage of total fluorescence positive cells (**D**), and the proportion of productively infected (GFP+, mKO2+) and latently infected (GFP−, mKO2+) cells (**E**), were determined using flow cytometry. (**F**,**G**), Jurkat cells were similarly pre-treated for 24 h with increasing concentrations of TmPyP4 and infected with HIV_{GKO} or $HIV_{GKO\text{-}DU3LTR}$ for 96 h in the presence of drug. The percentage of total fluorescence positive cells (**F**) and the proportion of productively infected (GFP+, mKO2+) and latently infected (GFP−, mKO2+) cells (**G**) were determined using flow cytometry. Data shown represent the average (+/− S.E.M.) from 3 independent experiments. Statistical analysis was performed using linear regression.

3.5. CPSF6 and LEDGF/p75 Influence Integration near Non-B DNA Motifs

Since LEDGF/p75 and CPSF6 promote integration into actively transcribed genes residing in gene-dense regions, we asked if CPSF6 and LEDGF/p75 influence the targeting of non-B DNA for integration [27,34,58,93–95]. We analyzed a previously well-characterized integration site dataset by Achuthan et al. (2018) who studied the impact of CPSF6 and LEDGF/p75 on integration site targeting [34]. In that study, CPSF6 function was inhibited by using the HIV-1 capsid mutant A77V, which impairs CPSF6 binding efficiency without severely decreasing infectivity. The LEDGF/p75 function was inhibited by treating the cells at the time of the infection with the allosteric integrase inhibitor BI-D, which competes with integrase-LEDGF/p75 binding. We generated the non-B DNA integration site profiles using this independently generated previously published dataset and asked if CPSF6 and/or LEDGF/p75 potentially influenced integration site targeting of non-B DNA motifs. Indeed, the cells defective for CPSF6 or LEDGF/p75 function exhibited significantly reduced integration within 500 bp of G4 DNA, mirror repeats, and short-tandem repeats compared to the wild type controls (Figure 5A, Tables S1 and S4). Analysis of the distribution of sites near non-B DNA features revealed an unexpected enrichment of integration sites 1–150 bp from G4 DNA in cells defective for the CPSF6 or LEDGF/p75 function compared to wild type cells (Figure 5A,B and Table S4). In addition, integration was more enriched 300–500 bp from slipped DNA in cells defective for CPSF6 and LEDGF/p75 function compared to wild type cells. Together, these data suggest that functional CPSF6 and LEDGF/p75 influence the distribution of the integration sites near several non-B DNA features.
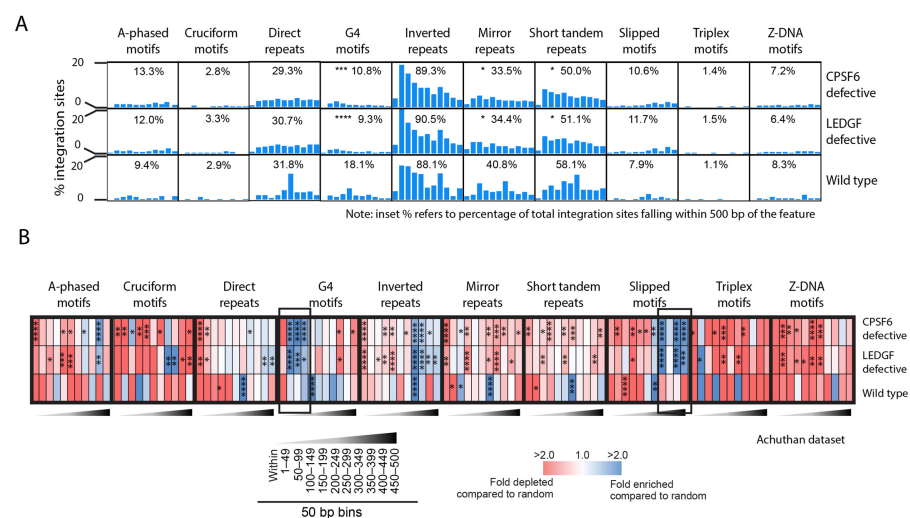

Figure 5. CPSF6 and LEDGF/p75 promote integration near several non-B DNA. (**A**), Graphs show the percentage of total unique HIV-1 integration sites located in or within 500 bp of various non-B

DNA motifs (distributed in 50 bp bins) from wild type (control), CPSF6 depleted, or LEDGF/p75 depleted cells. Inset numbers show the percentage of total unique integration sites falling within 500 bp of the non-B DNA motif. Statistical analysis is with respect to wild type. (**B**), Heatmap shows the fold enrichment (blue) or depletion (red) of integration sites at each distance interval from the non-B DNA motif compared to the random control. Black boxes highlight notable regions of enriched integration. Statistical analysis is with respect to the random control. Fisher's exact test; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

4. Discussion

The ability of HIV-1 to target non-B DNA structures for integration has several important implications for productive and latent infection, especially since genomic position effects have been shown to influence HIV-1 expression and latency reversal [18,20,30]. Several non-B DNA structures are associated with active genes in vivo and contribute to a dynamic interplay between DNA structure, chromatin organization, and transcriptional activities [67]. Non-B DNA structures can be recognized by non-B DNA-specific transcription factors, leading to transcriptional activation [96–99]. Conversely, the unusual non-B DNA structure can also block the binding of B-DNA-specific transcription factors, resulting in constitutive repression of adjacent genes and proviruses [43,50,51,89,100–102]. As such, the proximity of integrated proviruses to non-B DNA structures is likely to influence proviral expression, including reactivation potential by the LRAs. Indeed, we identified different integration site biases toward non-B DNA between latent and productive proviruses, especially those known to influence gene transcription (e.g., G4 DNA, triplex DNA, and Z-DNA).

The factors influencing the reactivation potential of latent proviruses are not fully understood. Latent proviruses that have a stronger response to LRAs tend to be located closer to active regulatory elements such as histone H3 acetylated at K27, monomethylated at K4, or trimethylated at K4 or K36, whereas non-reactivatable proviruses are biased toward heterochromatin with histone H3 trimethylated at K9 or K27 [18,20]. Non-B DNA structures, particularly G4 structures, have been shown to fine-tune chromatin organization within cells, although the mechanism underlying this activity is not fully understood [67]. Non-B DNA sequences are thought to alter the intrinsic sequence preference of nucleosomes, thereby affecting nucleosome occupancy [67,103]. Similarly, non-B DNA structures might sterically exclude nucleosomes, thereby affecting nucleosome positioning [67,104]. Indeed, G4 DNA has been shown to form in nucleosome-free regions in the genome [105]. Future research will help determine if the ability of G4 structures to locally and dynamically organize flanking nucleosomes also impacts the chromatin functional state near integrated HIV-1 proviruses, thereby influencing the transcriptional activity of those adjacent proviruses.

Other position effects such as the proximity of the integrated proviruses to enhancers can also impact proviral expression. Chen et al. (2017) showed that HIV-1 proviral expression is strongest when the proviruses are inserted close to endogenous enhancers, whereas latent insertions are mapped further away from enhancers [20]. We identified a strong correlation between latent proviruses that could be reactivated with LRAs and their integration within 50 bp of specific non-B DNA features that influence gene expression (e.g., G4 and Z-DNA). Indeed, enhancer–promoter interactions have been shown to be facilitated by G4 DNA [106]. It will be interesting to learn if G4 DNA located +/− 500 bp from latent proviral integration sites facilitates interactions with enhancers to render these proviruses more responsive to LRAs.

It has been previously suggested that an optimally tuned bias for integrating into transcriptionally active (euchromatin) versus inactive (heterochromatin) regions of the genome may help establish a diverse latent viral reservoir [107–109]. CPSF6 and LEDGF/p75 are two host proteins that promote integration into euchromatin [32,34,110–112]. Work performed by Vranckx et al. (2016) showed that by blocking LEDGF/p75–integrase interaction with the drug LEDGIN, integration site preferences shifted away from actively transcribed genes near the nuclear pores to the inner nuclear compartment [30]. Integration into tran-

scriptionally silent heterochromatin is often located close to the nucleolus in the inner core and LADs at the nuclear periphery [111]. It is widely known that heterochromatin is tightly wrapped by distinctive DNA-binding proteins (e.g., histones) and that this chromatin environment is generally unfavorable for integration. Interestingly, G4 structures are found in heterochromatin but are typically not located within the DNA wrapped around a histone octamer [89,105,113]. It is possible that this property creates a partially open state in the regions of heterochromatin, allowing an integration to occur near these structures.

Although the targeting of non-B DNA by pre-integration complexes likely involves multiple factors given the diversity of the non-B DNA, we identified LEDGF/p75 and CPSF6 from previously published integration site profiles as potential candidates that alter the targeting of some non-B DNA motifs. It could be that LEDGF/p75 and CPSF6 simply promote interactions between the pre-integration complexes and euchromatin where it then integrates near non-B DNA via another unknown mechanism mediated by, for example, certain non-B DNA binding proteins. Another possibility is that LEDGF/p75 and CPSF6 themselves can recognize specific non-B DNA structures to promote interactions between the pre-integration complex and the genomic DNA, leading to integration. More research is needed to investigate these and other possibilities.

The distribution of integration sites near non-B DNA was often heterogeneous and not located at precise nucleotide distances from the features. One explanation could be that the pre-integration complex is first attracted to the non-B DNA structure where it then integrates non-discriminately near the feature. Other explanations for the heterogeneity could be differences in the cellular transcriptional profiles that could affect the formation of non-B DNA structures, the adjacent nucleosome occupancy, and/or the DNA-binding proteins that generate steric constraints for an integration. Moreover, genetic polymorphisms (e.g., insertions/deletions) near non-B DNA motifs may also contribute to differences in the distance of the integration sites to the features in infected individuals. Interestingly, when we treated cells with low concentrations of G4 DNA stabilizing or destabilizing ligands, we observed increased integration in the windows of 150–200 bp (BRACO-19) and 250–500 bp (TmPyP4) around the motifs. It is intriguing that these distances correlate with the approximate distance of one, two, or three nucleosomes away from the G4 motif. The reason for this increase is currently unknown. One possible explanation is that at these low concentrations, the binding of the ligands perturbs the G4 structures and disrupts the G4 DNA- and/or nucleosome-binding proteins that typically block integration in these regions. Another possibility is that perturbation of these G4 DNA structures at these low concentrations leads to the temporary formation of alternative non-B DNA structures that promote integration. For example, it was shown by Masai et al. (2018) that some G4 DNA motifs may generate G4 structures that undergo a dynamic transition from one structure to another, with the potential to generate multiple G4-like structures on one or both DNA strands [114].

Previous searches to identify a consensus sequence for integration site targeting have only revealed an apparent weak palindromic sequence at the site of the insertion of several retroviruses [115–121]. Recent work by Kirk et al. (2016) challenged this notion by showing that the palindromic consensus sequence arises in the population average as a consequence of the non-palindromic motifs existing in equal proportions on the plus and minus strand of the target sequence [122]. Our study not only supports the notion that there is not likely a single palindromic consensus sequence at the integration site but shows that the integration site target sequences are heterogeneous in nature, many of which fall near non-B DNA structure-forming motifs. The inspection of each of the non-palindromic sequences identified by Kirk et al. (2016) in the subpopulations of the target integration sites from HTLV-1, HIV-1, MLV, ASLV, and PFV(IV) revealed that they all represent different non-B DNA sequences that are predicted to form slipped DNA structures. Moreover, the HTLV-1 'weak palindromic' consensus site sequence in Kirk et al.'s study that predominates from the population average is in fact a non-B DNA slipped motif [115,116,120,122]. This suggests that despite having distinct nucleotide consensus sequences (palindromic or non-

palindromic), the nature of the sequence is predicted to form a non-B DNA structure. Our findings using G4-stabilizing and -destabilizing ligands further highlights the likelihood that it may not be the primary DNA sequence itself that plays an important role in attracting the HIV-1 pre-integration complex, but rather it is the secondary structure formed by the non-B DNA motif itself that is important. Short-loop G4 motifs (e.g., $(TTAGGG)_n$) are highly enriched in the telomeres of the chromosomes, in which no strong bias for a HIV-1 integration has been observed. Our finding that HIV-1 demonstrated an integration bias for the long-loop G4 structures as opposed to the short-loop G4 structures raises the possibility that structural variations in some of these non-B DNA structures may also influence integration.

5. Conclusions

We have shown that different integration site biases for non-B DNA exist between productive and latent HIV-1 infection and correlate with the reactivation potential of latent proviruses. The manipulation of G4 DNA also altered the proportion of productively and latently infected cells in vitro. Moreover, we identified two key proteins involved in the integration process, LEDGF/p75 and CPSF6, as potential candidates that influence the distribution of integration sites near several non-B DNA motifs. Targeting non-B DNA structures directly as a form of antiretroviral therapy would be difficult given the abundance of non-B DNA in the human genome. However, a better understanding of how pre-integration complexes interact with non-B DNA may identify viral components critical for the selection of non-B DNA. These components could be targeted with novel small-molecule inhibitors to alter the proportion of productive and latent infections during an acute infection, and/or to enhance 'shock and kill' and 'block and lock' therapies in the future.

References

1. Arts, E.J.; Hazuda, D.J. HIV-1 Antiretroviral Drug Therapy. *Cold Spring Harb. Perspect. Med.* **2012**, *2*, a007161. [CrossRef] [PubMed]
2. Chun, T.W.; Finzi, D.; Margolick, J.; Chadwick, K.; Schwartz, D.; Siliciano, R.F. In Vivo Fate of HIV-1-Infected T Cells: Quantitative Analysis of the Transition to Stable Latency. *Nat. Med.* **1995**, *1*, 1284–1290. [CrossRef] [PubMed]
3. Chun, T.W.; Engel, D.; Berrey, M.M.; Shea, T.; Corey, L.; Fauci, A.S. Early Establishment of a Pool of Latently Infected, Resting CD4(+) T Cells during Primary HIV-1 Infection. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 8869–8873. [CrossRef] [PubMed]
4. Davey, R.T., Jr.; Bhat, N.; Yoder, C.; Chun, T.W.; Metcalf, J.A.; Dewar, R.; Natarajan, V.; Lempicki, R.A.; Adelsberger, J.W.; Miller, K.D.; et al. HIV-1 and T Cell Dynamics after Interruption of Highly Active Antiretroviral Therapy (HAART) in Patients with a History of Sustained Viral Suppression. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 15109–15114. [CrossRef]
5. Archin, N.M.; Liberty, A.L.; Kashuba, A.D.; Choudhary, S.K.; Kuruc, J.D.; Crooks, A.M.; Parker, D.C.; Anderson, E.M.; Kearney, M.F.; Strain, M.C.; et al. Administration of Vorinostat Disrupts HIV-1 Latency in Patients on Antiretroviral Therapy. *Nature* **2012**, *487*, 482–485. [CrossRef]
6. Deeks, S.G. HIV: Shock and Kill. *Nature* **2012**, *487*, 439–440. [CrossRef]
7. Finzi, D.; Blankson, J.; Siliciano, J.D.; Margolick, J.B.; Chadwick, K.; Pierson, T.; Smith, K.; Lisziewicz, J.; Lori, F.; Flexner, C.; et al. Latent Infection of CD4+ T Cells Provides a Mechanism for Lifelong Persistence of HIV-1, Even in Patients on Effective Combination Therapy. *Nat. Med.* **1999**, *5*, 512–517. [CrossRef]
8. Siliciano, J.D.; Kajdas, J.; Finzi, D.; Quinn, T.C.; Chadwick, K.; Margolick, J.B.; Kovacs, C.; Gange, S.J.; Siliciano, R.F. Long-Term Follow-up Studies Confirm the Stability of the Latent Reservoir for HIV-1 in Resting CD4+ T Cells. *Nat. Med.* **2003**, *9*, 727–728. [CrossRef]
9. Archin, N.M.; Vaidya, N.K.; Kuruc, J.D.; Liberty, A.L.; Wiegand, A.; Kearney, M.F.; Cohen, M.S.; Coffin, J.M.; Bosch, R.J.; Gay, C.L.; et al. Immediate Antiviral Therapy Appears to Restrict Resting CD4+ Cell HIV-1 Infection without Accelerating the Decay of Latent Infection. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 9523–9528. [CrossRef]
10. Rong, L.; Perelson, A.S. Modeling Latently Infected Cell Activation: Viral and Latent Reservoir Persistence, and Viral Blips in HIV-Infected Patients on Potent Therapy. *PLoS Comput. Biol.* **2009**, *5*, e1000533. [CrossRef]
11. Palmer, S.; Wiegand, A.P.; Maldarelli, F.; Bazmi, H.; Mican, J.M.; Polis, M.; Dewar, R.L.; Planta, A.; Liu, S.; Metcalf, J.A.; et al. New Real-Time Reverse Transcriptase-Initiated PCR Assay with Single-Copy Sensitivity for Human Immunodeficiency Virus Type 1 RNA in Plasma. *J. Clin. Microbiol.* **2003**, *41*, 4531–4536. [CrossRef] [PubMed]
12. Ruelas, D.S.; Greene, W.C. An Integrated Overview of HIV-1 Latency. *Cell* **2013**, *155*, 519–529. [CrossRef] [PubMed]
13. Dahabieh, M.S.; Battivelli, E.; Verdin, E. Understanding HIV Latency: The Road to an HIV Cure. *Annu. Rev. Med.* **2015**, *66*, 407–421. [CrossRef] [PubMed]
14. Hamer, D.H. Can HIV Be Cured? Mechanisms of HIV Persistence and Strategies to Combat It. *Curr. HIV Res.* **2004**, *2*, 99–111. [CrossRef] [PubMed]
15. Geeraert, L.; Kraus, G.; Pomerantz, R.J. Hide-and-Seek: The Challenge of Viral Persistence in HIV-1 Infection. *Annu. Rev. Med.* **2008**, *59*, 487–501. [CrossRef]
16. Savarino, A.; Mai, A.; Norelli, S.; El Daker, S.; Valente, S.; Rotili, D.; Altucci, L.; Palamara, A.T.; Garaci, E. "Shock and Kill" Effects of Class I-Selective Histone Deacetylase Inhibitors in Combination with the Glutathione Synthesis Inhibitor Buthionine Sulfoximine in Cell Line Models for HIV-1 Quiescence. *Retrovirology* **2009**, *6*, 52. [CrossRef]
17. Rasmussen, T.A.; Lewin, S.R. Shocking HIV out of Hiding: Where Are We with Clinical Trials of Latency Reversing Agents? *Curr. Opin. HIV AIDS* **2016**, *11*, 394–401. [CrossRef]
18. Battivelli, E.; Dahabieh, M.S.; Abdel-Mohsen, M.; Svensson, J.P.; Tojal Da Silva, I.; Cohn, L.B.; Gramatica, A.; Deeks, S.; Greene, W.C.; Pillai, S.K.; et al. Distinct Chromatin Functional States Correlate with HIV Latency Reactivation in Infected Primary CD4+ T Cells. *Elife* **2018**, *7*, e34655. [CrossRef]
19. Jordan, A.; Defechereux, P.; Verdin, E. The Site of HIV-1 Integration in the Human Genome Determines Basal Transcriptional Activity and Response to Tat Transactivation. *EMBO J.* **2001**, *20*, 1726–1738. [CrossRef]
20. Chen, H.-C.; Martinez, J.P.; Zorita, E.; Meyerhans, A.; Filion, G.J. Position Effects Influence HIV Latency Reversal. *Nat. Struct. Mol. Biol.* **2017**, *24*, 47–54. [CrossRef]
21. Sherrill-Mix, S.; Lewinski, M.K.; Famiglietti, M.; Bosque, A.; Malani, N.; Ocwieja, K.E.; Berry, C.C.; Looney, D.; Shan, L.; Agosto, L.M.; et al. HIV Latency and Integration Site Placement in Five Cell-Based Models. *Retrovirology* **2013**, *10*, 90. [CrossRef] [PubMed]
22. Dahabieh, M.S.; Ooms, M.; Brumme, C.; Taylor, J.; Harrigan, P.R.; Simon, V.; Sadowski, I. Direct Non-Productive HIV-1 Infection in a T-Cell Line Is Driven by Cellular Activation State and NFκB. *Retrovirology* **2014**, *11*, 17. [CrossRef] [PubMed]
23. Maldarelli, F.; Wu, X.; Su, L.; Simonetti, F.R.; Shao, W.; Hill, S.; Spindler, J.; Ferris, A.L.; Mellors, J.W.; Kearney, M.F.; et al. Specific HIV Integration Sites Are Linked to Clonal Expansion and Persistence of Infected Cells. *Science* **2014**, *345*, 179–183. [CrossRef] [PubMed]

24. Simonetti, F.R.; Sobolewski, M.D.; Fyne, E.; Shao, W.; Spindler, J.; Hattori, J.; Anderson, E.M.; Watters, S.A.; Hill, S.; Wu, X.; et al. Clonally Expanded CD4 + T Cells Can Produce Infectious HIV-1 in Vivo. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 1883–1888. [CrossRef] [PubMed]

25. Cohn, L.B.; Silva, I.T.; Oliveira, T.Y.; Rosales, R.A.; Parrish, E.H.; Learn, G.H.; Hahn, B.H.; Czartoski, J.L.; McElrath, M.J.; Lehmann, C.; et al. HIV-1 Integration Landscape during Latent and Active Infection. *Cell* **2015**, *160*, 420–432. [CrossRef]

26. Bruner, K.M.; Murray, A.J.; Pollack, R.A.; Soliman, M.G.; Laskey, S.B.; Capoferri, A.A.; Lai, J.; Strain, M.C.; Lada, S.M.; Hoh, R.; et al. Defective Proviruses Rapidly Accumulate during Acute HIV-1 Infection. *Nat. Med.* **2016**, *22*, 1043–1049. [CrossRef]

27. Ciuffi, A.; Llano, M.; Poeschla, E.; Hoffmann, C.; Leipzig, J.; Shinn, P.; Ecker, J.R.; Bushman, F. A Role for LEDGF/P75 in Targeting HIV DNA Integration. *Nat. Med.* **2005**, *11*, 1287–1289. [CrossRef]

28. Shun, M.-C.; Raghavendra, N.K.; Vandegraaff, N.; Daigle, J.E.; Hughes, S.; Kellam, P.; Cherepanov, P.; Engelman, A. LEDGF/P75 Functions Downstream from Preintegration Complex Formation to Effect Gene-Specific HIV-1 Integration. *Genes Dev.* **2007**, *21*, 1767–1778. [CrossRef]

29. Marshall, H.M.; Ronen, K.; Berry, C.; Llano, M.; Sutherland, H.; Saenz, D.; Bickmore, W.; Poeschla, E.; Bushman, F.D. Role of PSIP1/LEDGF/P75 in Lentiviral Infectivity and Integration Targeting. *PLoS ONE* **2007**, *2*, e1340. [CrossRef]

30. Vranckx, L.S.; Demeulemeester, J.; Saleh, S.; Boll, A.; Vansant, G.; Schrijvers, R.; Weydert, C.; Battivelli, E.; Verdin, E.; Cereseto, A.; et al. LEDGIN-Mediated Inhibition of Integrase–LEDGF/P75 Interaction Reduces Reactivation of Residual Latent HIV. *EBioMedicine* **2016**, *8*, 248–264. [CrossRef]

31. Singh, P.K.; Plumb, M.R.; Ferris, A.L.; Iben, J.R.; Wu, X.; Fadel, H.J.; Luke, B.T.; Esnault, C.; Poeschla, E.M.; Hughes, S.H.; et al. LEDGF/P75 Interacts with MRNA Splicing Factors and Targets HIV-1 Integration to Highly Spliced Genes. *Genes Dev.* **2015**, *29*, 2287–2297. [CrossRef] [PubMed]

32. Sowd, G.A.; Serrao, E.; Wang, H.; Wang, W.; Fadel, H.J.; Poeschla, E.M.; Engelman, A.N. A Critical Role for Alternative Polyadenylation Factor CPSF6 in Targeting HIV-1 Integration to Transcriptionally Active Chromatin. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E1054–E1063. [CrossRef] [PubMed]

33. Li, W.; Singh, P.K.; Sowd, G.A.; Bedwell, G.J.; Jang, S.; Achuthan, V.; Oleru, A.V.; Wong, D.; Fadel, H.J.; Lee, K.; et al. CPSF6-Dependent Targeting of Speckle-Associated Domains Distinguishes Primate from Nonprimate Lentiviral Integration. *mBio* **2020**, *11*, e02254-20. [CrossRef] [PubMed]

34. Achuthan, V.; Perreira, J.M.; Sowd, G.A.; Puray-Chavez, M.; McDougall, W.M.; Paulucci-Holthauzen, A.; Wu, X.; Fadel, H.J.; Poeschla, E.M.; Multani, A.S.; et al. Capsid-CPSF6 Interaction Licenses Nuclear HIV-1 Trafficking to Sites of Viral DNA Integration. *Cell Host Microbe* **2018**, *24*, 392–404.e8. [CrossRef] [PubMed]

35. Ho, Y.-C.; Shan, L.; Hosmane, N.N.; Wang, J.; Laskey, S.B.; Rosenbloom, D.I.S.; Lai, J.; Blankson, J.N.; Siliciano, J.D.; Siliciano, R.F. Replication-Competent Noninduced Proviruses in the Latent Reservoir Increase Barrier to HIV-1 Cure. *Cell* **2013**, *155*, 540–551. [CrossRef]

36. Jordan, A.; Bisgrove, D.; Verdin, E. HIV Reporducibly Establishes a Latent Infection after Acute Infection of T Cells Ni Vitro. *EMBO J.* **2003**, *22*, 1868–1877. [CrossRef]

37. Einkauf, K.B.; Lee, G.Q.; Gao, C.; Sharaf, R.; Sun, X.; Hua, S.; Chen, S.M.Y.M.; Jiang, C.; Lian, X.; Chowdhury, F.Z.; et al. Intact HIV-1 Proviruses Accumulate at Distinct Chromosomal Positions during Prolonged Antiretroviral Therapy. *J. Clin. Investig.* **2019**, *129*, 988–998. [CrossRef]

38. McAllister, R.G.; Liu, J.; Woods, M.W.; Tom, S.K.; Rupar, C.A.; Barr, S.D. Lentivector Integration Sites in Ependymal Cells from a Model of Metachromatic Leukodystrophy: Non-B DNA as a New Factor Influencing Integration. *Mol. Nucleic Acids* **2014**, *3*, e187. [CrossRef]

39. Bacolla, A.; Wells, R.D. Non-B DNA Conformations, Genomic Rearrangements, and Human Disease. *J. Biol. Chem.* **2004**, *279*, 47411–47414. [CrossRef]

40. Choi, J.; Majima, T. Conformational Changes of Non-BDNA. *Chem. Soc. Rev.* **2011**, *40*, 5893–5909. [CrossRef]

41. Wahls, W.P.; Wallace, L.J.; Moore, P.D. The Z-DNA Motif d(TG)30 Promotes Reception of Information during Gene Conversion Events While Stimulating Homologous Recombination in Human Cells in Culture. *Mol. Cell. Biol.* **1990**, *10*, 785–793. [PubMed]

42. 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* **2002**, *99*, 11593–11598. [CrossRef] [PubMed]

43. Verma, A.; Yadav, V.K.; Basundra, R.; Kumar, A.; Chowdhury, S. Evidence of Genome-Wide G4 DNA-Mediated Gene Expression in Human Cancer Cells. *Nucleic Acids Res* **2009**, *37*, 4194–4204. [CrossRef] [PubMed]

44. Waga, S.; Mizuno, S.; Yoshida, M. Chromosomal Protein HMG1 Removes the Transcriptional Block Caused by the Cruciform in Supercoiled DNA. *J. Biol. Chem.* **1990**, *265*, 19424–19428. [CrossRef]

45. Waga, S.; Mizuno, S.; Yoshida, M. Nonhistone Protein HMG1 Removes the Transcriptional Block Caused by Left-Handed Z-Form Segment in a Supercoiled DNA. *Biochem. Biophys. Res. Commun.* **1988**, *153*, 334–339. [CrossRef]

46. Jain, A.; Magistri, M.; Napoli, S.; Carbone, G.M.; Catapano, C.V. Mechanisms of Triplex DNA-Mediated Inhibition of Transcription Initiation in Cells. *Biochimie* **2010**, *92*, 317–320. [CrossRef]

47. Maher, L.J.; Dervan, P.B.; Wold, B. Analysis of Promoter-Specific Repression by Triple-Helical DNA Complexes in a Eukaryotic Cell-Free Transcription System. *Biochemistry* **1992**, *31*, 70–81. [CrossRef]

48. Bochman, M.L.; Paeschke, K.; Zakian, V.A. DNA Secondary Structures: Stability and Function of G-Quadruplex Structures. *Nat. Rev. Genet.* **2012**, *13*, 770–780. [CrossRef]

49. Brázda, V.; Laister, R.C.; Jagelská, E.B.; Arrowsmith, C. Cruciform Structures Are a Common DNA Feature Important for Regulating Biological Processes. *BMC Mol. Biol.* **2011**, *12*, 33. [CrossRef]

50. Delic, J.; Onclercq, R.; Moisan-Coppey, M. Inhibition and Enhancement of Eukaryotic Gene Expression by Potential Non-B DNA Sequences. *Biochem. Biophys. Res. Commun.* **1991**, *180*, 1273–1283. [CrossRef]

51. Tornaletti, S.; Park-Snyder, S.; Hanawalt, P.C. G4-Forming Sequences in the Non-Transcribed DNA Strand Pose Blocks to T7 RNA Polymerase and Mammalian RNA Polymerase II. *J. Biol. Chem.* **2008**, *283*, 12756–12762. [CrossRef] [PubMed]

52. Belotserkovskii, B.P.; De Silva, E.; Tornaletti, S.; Wang, G.; Vasquez, K.M.; Hanawalt, P.C. A Triplex-Forming Sequence from the Human c-MYC Promoter Interferes with DNA Transcription. *J. Biol. Chem.* **2007**, *282*, 32433–32441. [CrossRef] [PubMed]

53. Barr, S.D.; Ciuffi, A.; Leipzig, J.; Shinn, P.; Ecker, J.R.; Bushman, F.D. HIV Integration Site Selection: Targeting in Macrophages and the Effects of Different Routes of Viral Entry. *Mol. Ther.* **2006**, *14*, 218–225. [CrossRef] [PubMed]

54. Perrone, R.; Butovskaya, E.; Daelemans, D.; Palù, G.; Pannecouque, C.; Richter, S.N. Anti-HIV-1 Activity of the G-Quadruplex Ligand BRACO-19. *J. Antimicrob. Chemother.* **2014**, *69*, 3248–3258. [CrossRef] [PubMed]

55. Ofer, N.; Weisman-Shomer, P.; Shklover, J.; Fry, M. The Quadruplex r(CGG)n Destabilizing Cationic Porphyrin TMPyP4 Cooperates with HnRNPs to Increase the Translation Efficiency of Fragile X Premutation MRNA. *Nucleic Acids Res.* **2009**, *37*, 2712–2722. [CrossRef]

56. Morris, M.J.; Wingate, K.L.; Silwal, J.; Leeper, T.C.; Basu, S. The Porphyrin TmPyP4 Unfolds the Extremely Stable G-Quadruplex in MT3-MMP MRNA and Alleviates Its Repressive Effect to Enhance Translation in Eukaryotic Cells. *Nucleic Acids Res.* **2012**, *40*, 4137–4145. [CrossRef]

57. Ciuffi, A.; Barr, S.D. Identification of HIV Integration Sites in Infected Host Genomic DNA. *Methods* **2011**, *53*, 39–46. [CrossRef]

58. Schroder, A.; Shinn, P.; Chen, H.; Berry, C.; Ecker, J.R.; Bushman, F.D. HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots. *Cell* **2002**, *110*, 521–529. [CrossRef]

59. Quinlan, A.R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinform.* **2014**, *47*, 11.12.1–11.12.34. [CrossRef]

60. Cer, R.Z.; Donohue, D.E.; Mudunuri, U.S.; Temiz, N.A.; Loss, M.A.; Starner, N.J.; Halusa, G.N.; Volfovsky, N.; Yi, M.; Luke, B.T.; et al. Non-B DB v2.0: A Database of Predicted Non-B DNA-Forming Motifs and Its Associated Tools. *Nucleic Acids Res.* **2013**, *41*, D94–D100. [CrossRef]

61. Kazemier, H.G.; Paeschke, K.; Lansdorp, P.M. Guanine Quadruplex Monoclonal Antibody 1H6 Cross-Reacts with Restrained Thymidine-Rich Single Stranded DNA. *Nucleic Acids Res.* **2017**, *45*, 5913–5919. [CrossRef] [PubMed]

62. Henderson, A.; Wu, Y.; Huang, Y.C.; Chavez, E.A.; Platt, J.; Johnson, F.B.; Brosh, R.M.; Sen, D.; Lansdorp, P.M. Detection of G-Quadruplex DNA in Mammalian Cells. *Nucleic Acids Res.* **2014**, *42*, 860–869. [CrossRef] [PubMed]

63. Schindelin, J.; Arganda-Carreras, I.; Frise, E.; Kaynig, V.; Longair, M.; Pietzsch, T.; Preibisch, S.; Rueden, C.; Saalfeld, S.; Schmid, B.; et al. Fiji: An Open-Source Platform for Biological-Image Analysis. *Nat. Methods* **2012**, *9*, 676–682. [CrossRef] [PubMed]

64. Zhao, Y.; Du, Z.; Li, N. Extensive Selection for the Enrichment of G4 DNA Motifs in Transcriptional Regulatory Regions of Warm Blooded Animals. *FEBS Lett.* **2007**, *581*, 1951–1956. [CrossRef] [PubMed]

65. Du, Z.; Zhao, Y.; Li, N. Genome-Wide Analysis Reveals Regulatory Role of G4 DNA in Gene Transcription. *Genome Res.* **2008**, *18*, 233–241. [CrossRef]

66. Kouzine, F.; Wojtowicz, D.; Baranello, L.; Yamane, A.; Nelson, S.; Resch, W.; Kieffer-Kwon, K.-R.; Benham, C.J.; Casellas, R.; Przytycka, T.M.; et al. Permanganate/S1 Nuclease Footprinting Reveals Non-B DNA Structures with Regulatory Potential across a Mammalian Genome. *Cell Syst.* **2017**, *4*, 344–356.e7. [CrossRef]

67. Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nat. Chem.* **2013**, *5*, 182–186. [CrossRef]

68. Perrone, R.; Nadai, M.; Poe, J.A.; Frasson, I.; Palumbo, M.; Palù, G.; Smithgall, T.E.; Richter, S.N. Formation of a Unique Cluster of G-Quadruplex Structures in the HIV-1 Nef Coding Region: Implications for Antiviral Activity. *PLoS Med.* **2013**, *8*, e73121. [CrossRef]

69. Huppert, J.L. Structure, Location and Interactions of G-Quadruplexes. *FEBS J.* **2010**, *277*, 3452–3458. [CrossRef]

70. Weisman-Shomer, P.; Cohen, E.; Hershco, I.; Khateb, S.; Wolfovitz-Barchad, O.; Hurley, L.H.; Fry, M. The Cationic Porphyrin TMPyP4 Destabilizes the Tetraplex Form of the Fragile X Syndrome Expanded Sequence d(CGG)n. *Nucleic Acids Res.* **2003**, *31*, 3963–3970. [CrossRef]

71. Han, H.; Langley, D.R.; Rangan, A.; Hurley, L.H. Selective Interactions of Cationic Porphyrins with G-Quadruplex Structures. *J. Am. Chem. Soc.* **2001**, *123*, 8902–8913. [CrossRef] [PubMed]

72. Hu, M.-H.; Chen, S.-B.; Wang, B.; Ou, T.-M.; Gu, L.-Q.; Tan, J.-H.; Huang, Z.-S. Specific Targeting of Telomeric Multimeric G-Quadruplexes by a New Triaryl-Substituted Imidazole. *Nucleic Acids Res.* **2017**, *45*, 1606–1618. [CrossRef] [PubMed]

73. Grand, C.L.; Han, H.; Muñoz, R.M.; Weitman, S.; Von Hoff, D.D.; Hurley, L.H.; Bearss, D.J. The Cationic Porphyrin TMPyP4 Down-Regulates c-MYC and Human Telomerase Reverse Transcriptase Expression and Inhibits Tumor Growth in Vivo. *Mol. Cancer* **2002**, *1*, 565–573.

74. White, E.W.; Tanious, F.; Ismail, M.A.; Reszka, A.P.; Neidle, S.; Boykin, D.W.; Wilson, W.D. Structure-Specific Recognition of Quadruplex DNA by Organic Cations: Influence of Shape, Substituents and Charge. *Biophys. Chem.* **2007**, *126*, 140–153. [CrossRef] [PubMed]

75. Read, M.; Harrison, R.J.; Romagnoli, B.; Tanious, F.A.; Gowan, S.H.; Reszka, A.P.; Wilson, W.D.; Kelland, L.R.; Neidle, S. Structure-Based Design of Selective and Potent G Quadruplex-Mediated Telomerase Inhibitors. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4844–4849. [CrossRef]

76. Burger, A.M.; Dai, F.; Schultes, C.M.; Reszka, A.P.; Moore, M.J.; Double, J.A.; Neidle, S. The G-Quadruplex-Interactive Molecule BRACO-19 Inhibits Tumor Growth, Consistent with Telomere Targeting and Interference with Telomerase Function. *Cancer Res.* **2005**, *65*, 1489–1496. [CrossRef] [PubMed]

77. Tippana, R.; Hwang, H.; Opresko, P.L.; Bohr, V.A.; Myong, S. Single-Molecule Imaging Reveals a Common Mechanism Shared by G-Quadruplex-Resolving Helicases. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 8448–8453. [CrossRef]

78. Huppert, J.L.; Balasubramanian, S. G-Quadruplexes in Promoters throughout the Human Genome. *Nucleic Acids Res.* **2007**, *35*, 406–413. [CrossRef]

79. Maizels, N.; Gray, L.T.; Gellert, M.; Lipsett, M.N.; Davies, D.R.; Sen, D.; Gilbert, W.; Kim, J.; Cheong, C.; Moore, P.B.; et al. The G4 Genome. *PLoS Genet.* **2013**, *9*, e1003468. [CrossRef]

80. Eddy, J.; Vallur, A.C.; Varma, S.; Liu, H.; Reinhold, W.C.; Pommier, Y.; Maizels, N. G4 Motifs Correlate with Promoter-Proximal Transcriptional Pausing in Human Genes. *Nucleic Acids Res.* **2011**, *39*, 4975–4983. [CrossRef]

81. Du, Z.; Zhao, Y.; Li, N. Genome-Wide Colonization of Gene Regulatory Elements by G4 DNA Motifs. *Nucleic Acids Res.* **2009**, *37*, 6784–6798. [CrossRef] [PubMed]

82. Eddy, J.; Maizels, N. Selection for the G4 DNA Motif at the 5′ End of Human Genes. *Mol. Carcinog.* **2009**, *48*, 319–325. [CrossRef] [PubMed]

83. Kikin, O.; D'Antonio, L.; Bagga, P.S. QGRS Mapper: A Web-Based Server for Predicting G-Quadruplexes in Nucleotide Sequences. *Nucleic Acids Res.* **2006**, *34*, W676–W682. [CrossRef] [PubMed]

84. Scaria, V.; Hariharan, M.; Arora, A.; Maiti, S. Quadfinder: Server for Identification and Analysis of Quadruplex-Forming Motifs in Nucleotide Sequences. *Nucleic Acids Res.* **2006**, *34*, W683–W685. [CrossRef]

85. Bugaut, A.; Balasubramanian, S. A Sequence-Independent Study of the Influence of Short Loop Lengths on the Stability and Topology of Intramolecular DNA G-Quadruplexes. *Biochemistry* **2008**, *47*, 689–697. [CrossRef]

86. Frees, S.; Menendez, C.; Crum, M.; Bagga, P.S. QGRS-Conserve: A Computational Method for Discovering Evolutionarily Conserved G-Quadruplex Motifs. *Hum. Genom.* **2014**, *8*, 8. [CrossRef]

87. Puig Lombardi, E.; Londoño-Vallejo, A. A Guide to Computational Methods for G-Quadruplex Prediction. *Nucleic Acids Res.* **2020**, *48*, 1–15. [CrossRef]

88. Lago, S.; Tosoni, E.; Nadai, M.; Palumbo, M.; Richter, S.N. The Cellular Protein Nucleolin Preferentially Binds Long-Looped G-Quadruplex Nucleic Acids. *Biochim. Biophys. Acta (BBA)—Gen. Subj.* **2017**, *1861*, 1371–1381. [CrossRef]

89. Hoffmann, R.F.; Moshkin, Y.M.; Mouton, S.; Grzeschik, N.A.; Kalicharan, R.D.; Kuipers, J.; Wolters, A.H.G.; Nishida, K.; Romashchenko, A.V.; Postberg, J.; et al. Guanine Quadruplex Structures Localize to Heterochromatin. *Nucleic Acids Res.* **2015**, *44*, 152–163. [CrossRef]

90. Byrd, A.K.; Zybailov, B.L.; Maddukuri, L.; Gao, J.; Marecki, J.C.; Jaiswal, M.; Bell, M.R.; Griffin, W.C.; Reed, M.R.; Chib, S.; et al. Evidence That G-Quadruplex DNA Accumulates in the Cytoplasm and Participates in Stress Granule Assembly in Response to Oxidative Stress. *J. Biol. Chem.* **2016**, *291*, 18041–18057. [CrossRef]

91. Lyonnais, S.; Gorelick, R.J.; Mergny, J.L.; le Cam, E.; Mirambeau, G. G-Quartets Direct Assembly of HIV-1 Nucleocapsid Protein along Single-Stranded DNA. *Nucleic Acids Res.* **2003**, *31*, 5754–5763. [CrossRef] [PubMed]

92. Sundquist, W.I.; Heaphy, S. Evidence for Interstrand Quadruplex Formation in the Dimerization of Human Immunodeficiency Virus 1 Genomic RNA. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 3393–3397. [CrossRef] [PubMed]

93. Cherepanov, P.; Maertens, G.; Proost, P.; Devreese, B.; Van Beeumen, J.; Engelborghs, Y.; De Clercq, E.; Debyser, Z. HIV-1 Integrase Forms Stable Tetramers and Associates with LEDGF/P75 Protein in Human Cells. *J. Biol. Chem.* **2003**, *278*, 372–381. [CrossRef] [PubMed]

94. Maertens, G.; Cherepanov, P.; Pluymers, W.; Busschots, K.; De Clercq, E.; Debyser, Z.; Engelborghs, Y. LEDGF/P75 Is Essential for Nuclear and Chromosomal Targeting of HIV-1 Integrase in Human Cells. *J. Biol. Chem.* **2003**, *278*, 33528–33539. [CrossRef]

95. Lee, K.; Ambrose, Z.; Martin, T.D.; Oztop, I.; Mulky, A.; Julias, J.G.; Vandegraaff, N.; Baumann, J.G.; Wang, R.; Yuen, W.; et al. Flexible Use of Nuclear Import Pathways by HIV-1. *Cell Host Microbe* **2010**, *7*, 221–233. [CrossRef]

96. Brooks, T.A.; Hurley, L.H. The Role of Supercoiling in Transcriptional Control of MYC and Its Importance in Molecular Therapeutics. *Nat. Rev. Cancer* **2009**, *9*, 849–861. [CrossRef]

97. Cogoi, S.; Shchekotikhin, A.E.; Xodo, L.E. *HRAS* Is Silenced by Two Neighboring G-Quadruplexes and Activated by MAZ, a Zinc-Finger Transcription Factor with DNA Unfolding Property. *Nucleic Acids Res.* **2014**, *42*, 8379–8388. [CrossRef]

98. Kang, H.-J.; Le, T.V.T.; Kim, K.; Hur, J.; Kim, K.K.; Park, H.-J. Novel Interaction of the Z-DNA Binding Domain of Human ADAR1 with the Oncogenic c-Myc Promoter G-Quadruplex. *J. Mol. Biol.* **2014**, *426*, 2594–2604. [CrossRef]

99. Murat, P.; Balasubramanian, S. Existence and Consequences of G-Quadruplex Structures in DNA. *Curr. Opin. Genet. Dev.* **2014**, *25*, 22–29. [CrossRef]

100. Michelotti, G.A.; Michelotti, E.F.; Pullner, A.; Duncan, R.C.; Eick, D.; Levens, D. Multiple Single-Stranded Cis Elements Are Associated with Activated Chromatin of the Human c-Myc Gene in Vivo. *Mol. Cell Biol.* **1996**, *16*, 2656–2669. [CrossRef]

101. Lam, E.Y.N.; Beraldi, D.; Tannahill, D.; Balasubramanian, S. G-Quadruplex Structures Are Stable and Detectable in Human Genomic DNA. *Nat. Commun.* **2013**, *4*, 1796. [CrossRef] [PubMed]

102. Ray, B.K.; Dhar, S.; Henry, C.; Rich, A.; Ray, A. Epigenetic Regulation by Z-DNA Silencer Function Controls Cancer-Associated ADAM-12 Expression in Breast Cancer: Cross-Talk between MeCP2 and NF1 Transcription Factor Family. *Cancer Res.* **2013**, *73*, 736–744. [CrossRef] [PubMed]

103. Struhl, K.; Segal, E. Determinants of Nucleosome Positioning. *Nat. Struct. Mol. Biol.* **2013**, *20*, 267–273. [CrossRef]

104. Sadeh, R.; Allis, C.D. Genome-Wide "Re"-Modeling of Nucleosome Positions. *Cell* **2011**, *147*, 263–266. [CrossRef]

105. Wong, H.M.; Huppert, J.L. Stable G-Quadruplexes Are Found Outside Nucleosome-Bound Regions. *Mol. Biosyst.* **2009**, *5*, 1713–1719. [CrossRef] [PubMed]

106. Hegyi, H. Enhancer-Promoter Interaction Facilitated by Transiently Forming G-Quadruplexes. *Sci. Rep.* **2015**, *5*, 9165. [CrossRef] [PubMed]

107. Poeschla, E.M. Integrase, LEDGF/P75 and HIV Replication. *Cell. Mol. Life Sci.* **2008**, *65*, 1403–1424. [CrossRef]

108. Han, Y.; Wind-Rotolo, M.; Yang, H.-C.; Siliciano, J.D.; Siliciano, R.F. Experimental Approaches to the Study of HIV-1 Latency. *Nat. Rev. Microbiol.* **2007**, *5*, 95–106. [CrossRef]

109. Bisgrove, D.; Lewinski, M.; Bushman, F.; Verdin, E. Molecular Mechanisms of HIV-1 Proviral Latency. *Expert Rev. Anti-Infect.* **2005**, *3*, 805–814. [CrossRef]

110. Buffone, C.; Martinez-Lopez, A.; Fricke, T.; Opp, S.; Severgnini, M.; Cifola, I.; Petiti, L.; Frabetti, S.; Skorupka, K.; Zadrozny, K.K.; et al. Nup153 Unlocks the Nuclear Pore Complex for HIV-1 Nuclear Translocation in Nondividing Cells. *J. Virol.* **2018**, *92*, e00648-18. [CrossRef]

111. Marini, B.; Kertesz-Farkas, A.; Ali, H.; Lucic, B.; Lisek, K.; Manganaro, L.; Pongor, S.; Luzzati, R.; Recchia, A.; Mavilio, F.; et al. Nuclear Architecture Dictates HIV-1 Integration Site Selection. *Nature* **2015**, *521*, 227–231. [CrossRef] [PubMed]

112. Lelek, M.; Casartelli, N.; Pellin, D.; Rizzi, E.; Souque, P.; Severgnini, M.; Di Serio, C.; Fricke, T.; Diaz-Griffero, F.; Zimmer, C.; et al. Chromatin Organization at the Nuclear Pore Favours HIV Replication. *Nat. Commun.* **2015**, *6*, 6483. [CrossRef] [PubMed]

113. Van Holde, K.; Zlatanova, J. Unusual DNA Structures, Chromatin and Transcription. *Bioessays* **1994**, *16*, 59–68. [CrossRef] [PubMed]

114. Masai, H.; Kakusho, N.; Fukatsu, R.; Ma, Y.; Kanoh, Y.; Nagasawa, K. Molecular Architecture of G-Quadruplex Structures Generated on Duplex Rif1-Binding Sequences. *J. Biol. Chem.* **2018**, *293*, 17033–17049. [CrossRef] [PubMed]

115. Wu, X.; Li, Y.; Crise, B.; Burgess, S.M.; Munroe, D.J. Weak Palindromic Consensus Sequences Are a Common Feature Found at the Integration Target Sites of Many Retroviruses. *J. Virol.* **2005**, *79*, 5211–5214. [CrossRef]

116. Holman, A.G.; Coffin, J.M. Symmetrical Base Preferences Surrounding HIV-1, Avian Sarcoma/Leukosis Virus, and Murine Leukemia Virus Integration Sites. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6103–6107. [CrossRef]

117. Carteau, S.; Hoffmann, C.; Bushman, F.D. Chromosome Structure and HIV-1 CDNA Integration: Centromeric Alphoid Repeats Are a Disfavored Target. *J. Virol.* **1998**, *72*, 4005–4014. [CrossRef]

118. Stevens, S.W.; Griffith, J.D. Sequence Analysis of the Human DNA Flanking Sites of Human Immunodeficiency Virus Type 1 Integration. *J. Virol.* **1996**, *70*, 6459–6462. [CrossRef]

119. Wang, G.P.; Ciuffi, A.; Leipzig, J.; Berry, C.C.; Bushman, F.D. HIV Integration Site Selection: Analysis by Massively Parallel Pyrosequencing Reveals Association with Epigenetic Modifications. *Genome Res.* **2007**, *17*, 1186–1194. [CrossRef]

120. Derse, D.; Crise, B.; Li, Y.; Princler, G.; Lum, N.; Stewart, C.; McGrath, C.F.; Hughes, S.H.; Munroe, D.J.; Wu, X. Human T-Cell Leukemia Virus Type 1 Integration Target Sites in the Human Genome: Comparison with Those of Other Retroviruses. *J. Virol.* **2007**, *81*, 6731–6741. [CrossRef]

121. Berry, C.; Hannenhalli, S.; Leipzig, J.; Bushman, F.D. Selection of Target Sites for Mobile DNA Integration in the Human Genome. *PLoS Comput. Biol.* **2006**, *2*, e157. [CrossRef] [PubMed]

122. Kirk, P.D.W.; Huvet, M.; Melamed, A.; Maertens, G.N.; Bangham, C.R.M. Retroviruses Integrate into a Shared, Non-Palindromic DNA Motif. *Nat. Microbiol.* **2016**, *2*, 16212. [CrossRef] [PubMed]