
White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/243553/

Version: Published Version

Article:
Monti, Ludovica, Firth, Genta, Correia Faria, Joana Raquel et al. (2026) Genome-wide mapping of DNA G-quadruplexes in Trypanosoma brucei chromatin reveals enrichment in coding regions and transcription start sites. Nucleic Acids Research. gkag637. ISSN: 0305-1048

https://doi.org/10.1093/nar/gkag637

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial (CC BY-NC) licence. This licence allows you to remix, tweak, and build upon this work non-commercially, and any new works must also acknowledge the authors and be non-commercial. You don't have to license any derivative works on the same terms. More information and the full terms of the licence here:
https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.


UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Genome-wide mapping of DNA G-quadruplexes in *Trypanosoma brucei* chromatin reveals enrichment in coding regions and transcription start sites

Ludovica Monti, Genta Firth, Joana R Correia Faria, John M Kelly, Gem Flint, Silvia Galli, Thomas E Maher, Htay Mon Aye, Marco Di Antonio

Downloaded from https://academic.oup.com/nar/article/54/12/gkag637/8721998 by University of York user on 17 July 2026

Nucleic Acids Research, 2026, **54**, gkag637
https://doi.org/10.1093/nar/gkag637
Genomics

Genome-wide mapping of DNA G-quadruplexes in *Trypanosoma brucei* chromatin reveals enrichment in coding regions and transcription start sites

Ludovica Monti[1,2,*], Genta Firth[1], Joana R. Correia Faria[3,4], John M. Kelly[5], Gem Flint[2,6], Silvia Galli[2], Thomas E. Maher[2,6], Htay Mon Aye[3,4], Marco Di Antonio[2,6,7,*]

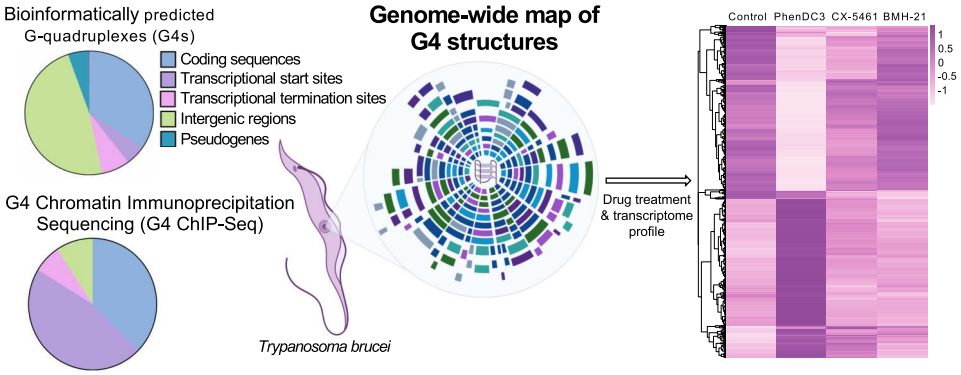
[1]School of Chemistry and The Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, Victoria 3052, Australia
[2]Department of Chemistry, Molecular Sciences Research Hub, Imperial College London, 82 Wood Lane, London W12 0BZ, United Kingdom
[3]Department of Biology, University of York, Heslington, York YO10 5DD, United Kingdom
[4]York Biomedical Research Institute, University of York, Heslington, York YO10 5DD, United Kingdom
[5]Department of Infection Biology, London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom
[6]Institute of Chemical Biology, Molecular Sciences Research Hub, Imperial College London, 82 Wood Lane, London W12 0BZ, United Kingdom
[7]The Francis Crick Institute, 1 Midland Road, London NW1 1AT, United Kingdom

*To whom correspondence should be addressed. Email: m.di-antonio@imperial.ac.uk
Correspondence may also be addressed to Ludovica Monti. Email: ludovica.monti@unimelb.edu.au

Abstract

G-quadruplexes (G4s) are non-canonical DNA structures formed in guanine-rich sequences that are proposed to act as regulatory elements in trypanosomatid parasites, including *Trypanosoma brucei*, the causative agent of African sleeping sickness. However, their functional roles remain poorly understood, largely due to limited knowledge of their genomic distribution. Herein, we performed *in silico* analyses across 64 trypanosomatid species, uncovering a high degree of variability in G4 prevalence and species-specific patterns. We generated the first chromatin-based, genome-wide G4 map in *T. brucei* using G4 chromatin immunoprecipitation followed by sequencing (G4 ChIP-Seq), revealing enrichment within gene-associated regions, including coding DNA sequences (CDSs), and transcription boundaries such as transcription start sites (TSSs) and transcription termination sites. This pattern diverges markedly from previous genome-wide G4 maps in humans, suggesting that G4s may play roles unique to trypanosome biology. To investigate their functional relevance, we profiled the transcriptome of *T. brucei* upon treatment with the G4-stabilizing ligand PhenDC3. PhenDC3 induced targeted transcriptional perturbation of genes bearing G4s, particularly those located within CDSs and TSSs. Altogether, our findings highlight a distinctive role for G4s in regulating gene expression in *T. brucei* and support their potential as therapeutic targets in the treatment of African sleeping sickness.

Graphical abstract



Received: July 23, 2025. **Revised:** May 13, 2026. **Accepted:** May 16, 2026
© The Author(s) 2026. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Downloaded from https://academic.oup.com/nar/article/54/12/gkag637/8721998 by University of York user on 17 July 2026

Introduction

Human African trypanosomiasis, commonly known as 'sleeping sickness,' is a parasitic disease caused by protozoa of the *Trypanosoma* genus, predominantly transmitted through the bite of infected tsetse flies. This disease is classified as one of the neglected tropical diseases (NTDs), significantly affecting impoverished populations in sub-Saharan Africa [1, 2]. An estimated 55 million people in 36 sub-Saharan African countries are at risk, with ∼1000 cases reported annually [1]. The disease often progresses to late-stage infections that affect the central nervous system, causing neurological damage, and is fatal without timely treatment [3]. Current treatments for African trypanosomiasis present significant challenges, including toxicity, ranging from severe systemic side effects to treatment-related mortality (e.g. fatal reactive encephalopathy following melarsoprol treatment), as well as complex administration requiring prolonged intravenous infusion and the need for hospitalization [4, 5]. The emergence of drug-resistant strains poses further risks, especially in rural areas where access to healthcare is limited [6–8]. Fexinidazole, introduced in 2019 as an oral treatment, marked a significant advancement in simplifying treatment protocols [9]. However, its efficacy in advanced-stage disease, as well as in children below a certain age (<6 years) and weight (<20 kg), pregnant women, and breastfeeding mothers remains unclear [10, 11].

The impact of African trypanosomiasis extends beyond human health, affecting agriculture and economic stability across the region [12]. Animal trypanosomiasis, known as Nagana, has severe consequences for the livestock industry, particularly cattle, where weight loss, reduced productivity, and high mortality lead to economic losses in agriculture exceeding $1 billion annually [13, 14]. Current therapeutic options include diminazene aceturate, isometamidium chloride, and homidium bromide/chloride; however, their narrow therapeutic index and the widespread emergence of drug resistance represent growing concerns, highlighting the urgent need for innovative therapeutic strategies [8, 15]. Identifying new molecular targets is therefore essential to addressing these challenges and advancing the development of safer, more effective treatments against drug-resistant strains.

G-quadruplexes (G4s) are DNA secondary structures formed in guanine-rich sequences through self-assembly mediated by Hoogsteen hydrogen bonding [16]. These structures consist of stacked guanine tetrads, which are stabilized by the presence of central monovalent cations such as K^+ or Na^+ [17, 18]. G4 formation is typically observed in genomic sequences containing at least three guanine tetrads separated by loops of up to seven nucleotides (GGG-$N_{1–7}$-GGG-$N_{1–7}$-GGG-$N_{1–7}$) [19]. However, recent experimental evidence has revealed that sequences deviating from this general consensus motif can also form stable G4s under physiological conditions, which are often referred to as non-canonical G4s [20]. G4s are non-randomly distributed across the genomes of all eukaryotic organisms. In humans, predictive algorithms [21] and next-generation sequencing techniques [20, 22] have revealed enrichment of G4s in regulatory regions, such as telomeres and promoters of proto-oncogenes [23] (i.e. *c-MYC* [24] and *c-KIT* [25]). These findings underscore the potential relevance of G4s in key cellular processes, including transcriptional regulation, replication, and genome stability [18, 26]. The development of genome-wide mapping strategies to underpin the distribution of G4s in chromatin has further revealed how the formation of these structures has a direct impact on promoting transcription, especially when formed at promoters [27, 28]. More recently, we identified that clusters of G4s formed at intergenic and intronic regions can act as super enhancers (super-G4s), further highlighting a potential role of G4s as epigenetic markers for transcriptional control [29]. Based on these observations, significant efforts have been made in the development of G4 targeting drugs [30, 31], with two candidates—CX-5461 and QN-302—progressing through clinical trials for the treatment of tumour malignancies [32–34], underscoring the translational impact of G4-targeting therapeutic approaches. This has promoted the investigation of G4-targeting strategiesbeyond cancer treatment and across various organisms, including pathogens (e.g. viruses, bacteria, and parasites) [35–37], and others (e.g. yeasts, helminths, and plants) [22, 38, 39]. For example, in protozoan parasites such as *Plasmodium falciparum* (the causative agent of malaria) and *T. brucei*, putative G4s have been identified in the surrounding regions of virulence genes and trypanosomatid-specific epigenetic modifications (i.e. *base J*) linked to parasite antigenic variation and survival in their hosts [35, 39–41]. This has prompted the development of a series of new G4-ligands with antiproliferative effects in *T. brucei* cells, highlighting the potential of G4s as therapeutic targets against *Trypanosoma* infections [42–52]. Despite these promising findings, underpinning the molecular mechanism that justifies the effectiveness of G4-ligands to treat *Trypanosoma* infections remains challenging due to the absence of a reliable genomic map of G4 distribution in this parasite. This limits our ability to fully harness the therapeutic potential of these DNA structures for treating sleeping sickness and other trypanosomatid infections.

In this study, we present the first genome-wide mapping of G-quadruplex (G4) structures in *T. brucei* chromatin and explore their functional relevance. Through G4-targeting chromatin immunoprecipitation followed by sequencing (G4 ChIP-Seq) and transcriptomic profiling, we reveal that G4s are widely distributed across the *T. brucei* genome, with a notable enrichment in gene-associated regions, including coding sequences (CDSs), and transcription boundary regions corresponding to transcription start sites (TSSs) and transcription termination sites (TTSs). This distribution differs from the promoter-centric enrichment commonly observed in mammals. Functional assays using the G4-stabilizing ligand PhenDC3 showed significant gene expression perturbation of G4-containing genes, particularly within TSSs (upregulation) and TTSs (downregulation), suggesting a distinct mechanism of G4-mediated gene regulation in *T. brucei*. These findings underscore the potential of targeting G4 structures to perturb essential gene networks in trypanosomes and inform novel strategies for therapeutic intervention.

Materials and methods

In silico bioinformatic analyses

Genome sequences. A total of 64 genomes representing seven trypanosomatid genera were downloaded from the NCBI genome database: *Trypanosoma*, *Leishmania*, *Crithidia*, *Phytomonas*, *Herpetomonas*, *Leptomonas*, and *Porcisia*. Additionally, three non-trypanosomatid genomes were included for comparative analyses: two *Schistosoma* genomes [53] (a

Downloaded from https://academic.oup.com/nar/article/54/12/gkag637/8721998 by University of York user on 17 July 2026

trematode parasite that causes a neglected tropical disease) and the human genome [54] for comparison. Accession numbers for all genomes are provided in Supplementary Table S1.

G4Hunter analysis. The G4Hunter algorithm [55–57] was used to identify putative G-quadruplex sequences (PQSs) across multiple genomes. A custom Python script was developed to automate the predictive analysis in iterative loops. The analysis utilized two key input parameters: window size and threshold. The window size defines the length of DNA fragments analysed, while the threshold determines the propensity of a sequence to form a G4 structure. Unless otherwise specified, the default G4Hunter settings of 25-nucleotide window size and a threshold of 1.2 were used, as these parameters have been previously validated to achieve a balance between minimizing false positives and false negatives [53, 55]. To further refine the analysis, additional evaluations were conducted at higher thresholds of 1.5, 1.8, and 2.0. The frequency of PQSs per 1000 bp for each species was determined by dividing the number of PQSs identified using G4Hunter by the corresponding genome length and multiplying the result by 1000.

Cell cultures

Trypanosoma brucei. *T. brucei* Lister 427 bloodstream forms were grown in HMI-11 medium (Gibco) supplemented with 10% foetal bovine serum (FBS, Sigma–Aldrich) at 37°C and 5% CO_2. The density of cell cultures was maintained below 1.5×10^6 cells/ml and sub-passaged every 24–48 h.

Human and monkey cells. HeLa cells were grown in culture flasks or microtiter plates after initial seeding at 1×10^4 cells/ml in DMEM medium (Sigma–Aldrich) supplemented with 5% FBS. Cells were maintained at 37°C in a humidified atmosphere containing 5% CO_2 and sub-cultured every 5 days. Human colon carcinoma COLO 205 (COLO) cells and African green monkey foetal kidney epithelial (MA-104) cells were cultured in MEM medium (Sigma–Aldrich) supplemented with 5% FBS and maintained at 37°C in a humidified atmosphere containing 5% CO_2.

Preparation of recombinant BG4 antibody

Recombinant BG4 was expressed as previously described [58]. Briefly, BL21(DE3) *E. coli* cells were inoculated with the expression vector pSANG10-3FBG4 (Addgene, plasmid no. 55756) [59] via heat shock, then grown overnight in 2×TY medium containing glucose (2%) and kanamycin (50 mg/ml, Gibco) before expanding into autoinduction medium. The culture was grown at 37°C for 6 h, followed by overnight incubation at 25°C. After harvesting the cells, protein purification was performed using nickel affinity chromatography, elution in the presence of imidazole, followed by dialysis in PBS overnight. Protein purity and concentration were assessed by sodium dodecyl sulphate–polyacrylamide gel electrophoresis, and antibody functionality was evaluated by ELISA [60].

G4-targeting chromatin immunoprecipitation followed by sequencing

Cell lysis and chromatin fragmentation were carried out as in ref. [61, 62]. Briefly, 2×10^8 parasites were cross-linked with 1% formaldehyde for 20 min at room temperature. Chromatin was sonicated using a Bioruptor (Diagenode) with sonication beads (Diagenode) for 20 cycles (30 s on, 30 s off) using the high setting. Cell debris was pelleted by centrifugation, and 10% Triton-X was added to each sample. Assessment of successful chromatin sonication to a required fragment size distribution of 100–500 bp was carried out using either 2% (wt/vol) agarose gel electrophoresis or the Agilent 4200 TapeStation system (version 4.1.1) and the Agilent D5000 screentape following manufacturer's instructions. Aliquots of chromatin were flash frozen in liquid nitrogen and stored at −80°C until further use.

Chromatin immunoprecipitation (ChIP) was carried out as in [60] with the following modifications. Chromatin ([59–76] ng/μl) was thawed on ice before being supplemented with Triton X-100 (2%) and incubated at room temperature for 10 min. Chromatin (9.4 μl/sample; 720 ng/sample), RNase A (1 μl/sample, Ambion™), and blocking buffer [37 μl/sample; containing intracellular salt buffer with 1% bovine serum albumin (BSA) in Milli-Q water, filter-sterilized before use] were incubated on a thermomixer under rotation (1400 rpm) at 37°C for 20 min. BG4 (0.75 μl/sample; 136 ng/sample) was added to the ChIP samples. One sample was left untreated with BG4 and kept on ice to serve as the input control sample. The ChIP samples treated with BG4 were incubated at 16°C with 1400 rpm rotation for 1 h. Meanwhile, Anti-FLAG antibodies (5 μl/sample, Millipore) were washed three times with blocking buffer (50 μl/sample) using the DynaMag-2 magnet and then resuspended in blocking buffer (50 μl/sample). Anti-FLAG bead suspension (50 μl/sample) was added to the ChIP samples and incubated at 16°C with 1400 rpm rotation for 1 h. The ChIP samples were then washed three times at 4°C with ice-cold wash buffer [200 μl/sample; containing: 100 mM KCl, 0.1% (vol/vol) TWEEN-20 and 10 mM Tris, pH 7.4, in Milli-Q water, and filter-sterilized before use]. Two additional washes were performed using wash buffer (200 μl/sample), followed by a 10-min incubation at 37°C, and 1400 rpm TE buffer was then added to all samples (including the input) to a final volume of 75 μl, followed by the addition of 1 μl of proteinase K (20 mg/ml, Invitrogen). The samples were incubated for 1 h at 37°C with continuous rotation at 1400 rpm, followed by a 2-h incubation at 65°C with 1400 rpm rotation. Samples were then purified using MinElute Reaction Cleanup Kit, following the manufacturer's instructions, with elution in 25 μl of elution buffer. The DNA concentration of the samples was measured using a Qubit fluorometer with the Qubit dsDNA HS Kit (Thermo Fisher Scientific, MA, USA). Chromatin was derived from two independent *T. brucei* cultures (representing two biological replicates). For each biological replicate, two to three technical ChIP replicates were performed in parallel, together with a corresponding input control sample.

Prior to sequencing, quantitative PCR (qPCR) was used to determine enrichment of G4-forming regions over non-G4-forming regions in ChIP samples compared with input sample. Purified DNA samples were diluted in DNase/RNase-free water, and 5 μl of the diluted DNA was amplified via qPCR using 10 μl of Fast SYBR Green Master Mix and 2.5 μl each of forward and reverse primers at a final concentration of 1 μM (primer sequences listed in Table 1). Technical duplicates were tested for each primer set in both ChIP and input samples. qPCR was performed on a Stratagene Mx3005P Real-Time PCR system (Agilent Technologies, CA, USA) using the following program of initial denaturing at 95°C for 20 s, followed by 40 cycles of denaturing at 95°C for 3 s, and annealing and extension at 60°C for 30 s.

To calculate the fraction of recovery (% Input), Ct values were adjusted relative to the input sample using the formula: $2^{(\text{Ct value of the Input} - \text{Ct value of the ChIP})} \times 100$. To estimate the G4-

Downloaded from https://academic.oup.com/nar/article/54/12/gkag637/8721998 by University of York user on 17 July 2026

Table 1. Forward and reverse primers used for qPCR to assess G-quadruplex (G4) enrichment at G4-positive and G4-negative loci within the bloodstream expression site 1 (BES1) of *T. brucei*

Primer name	Genomic region	Start	End	Sequence	G4-region	Melting temperature (°C)	G4Hunter PQS (1.2 threshold, 25 nt)
BES1(+)-forward	BES1	69 864	70 446	5′-ACCGACACACCAACCAAC-3′	+	62	+
BES1(+)-reverse	BES1			5′-TCGACGTTGCATTCTGCTATAA-3′	+	62	
BES1(−)-forward	BES1	62 586	62 772	5′-CGTATTTGATATCTCCTTCGTCACTA-3′	−	62	−
BES1(−)-reverse	BES1			5′-CTGGATGTTACGCCATATGAATAC-3′	−	61	

fold enrichment, the % Input of a G4-positive region was divided by the % Input of a G4-negative region. Fold enrichment was then calculated by dividing the average percentage input of G4 regions by that of the non-G4 regions.

Libraries were prepared from input and immunoprecipitated DNA using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, MA, USA). Final libraries were sequenced across two runs on the Illumina NovaSeq 6000 at either 200 or 300 cycles of paired-end reads at a depth of 15 and 38 million reads per sample.

ChIP-Seq data analysis

Recombinant read quality was assessed using *FastQC* (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and adapter sequences were trimmed with *fastp* [63]. Trimmed reads were aligned to the 2018 genome assembly of the Lister 427 version 10 [64]. (Tb427v10) extracted from TriTrypDB, using *Bowtie2* [65] with default settings. Reads were filtered and sorted using *SAMtools* [66] with the following arguments: *-F 2308* and *-q 10* to exclude unaligned, secondary, and supplementary reads and reads with a mapping quality (MAPQ) < 10. Duplicate reads were removed using *Picard MarkDuplicates* (https://broadinstitute.github.io/picard/) with the *REMOVE_DUPLICATES = TRUE* option or using *SAMtools* [66]. In the SAMtools workflow, properly paired reads were first filtered (*samtools view -bh -f 0×2*), followed by sorting (*samtools sort*), mate information correction (*samtools fixmate*), a second sorting step, and duplicate marking using *samtools markdup*. *Bowtie2* alignment files were analysed for peak calling (i.e. read enrichment) using *MACS2* [67] with a user-defined genome size of -g 2.67e7; 200 nt fragment size for shifting model (–nomodel –extsize 200); and minimum false discovery rate: ≥0.01 (-q 0.01). Log_2 ratio of ChIP signal to input signal was generated using *bamCompare* from the *deepTools* [68] suite (version 2.4.2) with a bin size of 50 bp. The resulting bigWig files were visualized using the Integrative Genomics Viewer (IGV, version 2.19.4) [69]. *Bowtie2* and *MACS2* output files were also visualized with IGV. Common and unique peaks between replicates were identified with *bedtools intersect* [70].

Genomic feature annotation and enrichment analysis

Genomic distribution of ChIP-Seq peaks was annotated using the *GenomicRanges* package in R [71]. Published G4-seq peak coordinates for *T. brucei* under K^+ and K^+/PDS G4-stabilizing conditions from [22] (GEO accession numbers GSM3003555 and GSM3003556) were downloaded and combined across strands into unified BED files. Coordinates were then converted to the *T. brucei* Lister 427 v10 genome using *Liftoff* [72] following BED-to-GFF conversion, and successfully mapped features were merged. This workflow was implemented using Nextflow [73]. G4 ChIP-Seq and G4-seq peaks were assigned to genomic features based on the extent of their overlap with annotated regions. Gene feature annotations were based on the *T. brucei* 427 genome assembly Tb427v10 [64]. To supplement this, 341 of the 342 annotated TSSs and TTSs from the latest genome assembly (v12) [74] were successfully mapped to the v10 assembly using *Liftoff* [72]. Intergenic regions were defined as genomic intervals not overlapping any annotated gene feature (i.e. protein-coding genes, TSS, TTS, pseudogenes, and ncRNA) and were included as a separate feature category in the annotation and enrichment analysis.

To assess genomic enrichment, random genome-wide shuffling of G4 ChIP-Seq peaks generated in this study and G4-seq peaks from [22] was performed three independent times using *bedtools shuffle* (v2.31.1) [70], and the overlap was assessed in each randomized case. Fold enrichment was calculated for each genomic feature by dividing the observed overlap of ChIP-Seq peaks by the overlap from each of three randomized datasets, producing three-fold enrichment values. The standard deviation (SD) of these values was then computed. For polycistronic transcription unit (PTU) analysis, gene-to-PTU associations were constructed using genome annotations from Tb427v12. A gene–PTU mapping dictionary was generated in R using the *GenomicRanges* package [71], allowing each gene to be linked to its corresponding PTU for downstream enrichment and localization analysis.

Radial plots for the G4 ChIP-Seq data obtained from this study and G4-seq data from [22] were generated using *Circos* hosted on Galaxy [75, 76] and edited using *Affinity Designer 2* (https://affinity.serif.com). The genomic regions were based on coordinates defined by the official genome FASTA assembly, excluding unitig and unplaced-contig fragments. Bloodstream expression sites (BESs) were treated as a distinct track due to their specialized and independently regulated nature within the subtelomeric domains. The log_2 fold change (ChIP/Input) across the genome was calculated using *bamCompare* from the *deepTools* [68] suite with a bin size of 1000 bp, generating bigWig files for each ChIP replicate. Replicated bigWig files were then averaged using *bigwigAverage* function from *deepTools* [68], and the resulting mean bigWig file was converted to WIG format using *bigWigToWig* via Galaxy [77]. Published G4-seq peak coordinates for *T. brucei* under K^+ and K^+/PDS G4-stabilizing conditions from [22] were downloaded and processed as described under '*Genomic feature*

Downloaded from https://academic.oup.com/nar/article/54/12/gkag637/8721998 by University of York user on 17 July 2026

annotation and enrichment analysis.' G4 peak scores on the negative strand were assigned negative values to generate the final peak set.

Cell viability assays with G4-ligands

Antitrypanosomal activity against *T. brucei* Lister 427 bloodstream forms and cytotoxicity in HeLa, COLO, and MA-104 cells were assessed for selected G4-targeting molecules. BMH-21 and CX-5461 were purchased from Cayman Chemical (Item No. 22282, MI, USA) and Selleckchem.com (Item No. S2684, TX, USA), respectively. PhenDC3 was synthesized in the lab following established procedures [78]. The viability of *T. brucei* and HeLa, COLO, and MA-104 cells was evaluated in 96-well microtiter plates (200 µl per well). Each viability assay was initiated with 2.5×10^4 parasites/ml or 1×10^4 HeLa/COLO/MA-104 cells/ml, followed by the addition of test compounds at concentrations ranging from nM to µM. For *T. brucei,* plates were incubated for 48 h, after which resazurin (20 µl at 0.125 mg/ml) was added, and the plates were incubated for an additional 16 h. For HeLa, COLO, and MA-104 cells, plates were incubated for 5 days, resazurin was added (as above), and the plates were incubated for another 6 h. Fluorescence was measured using a BMG FLUOstar Omega plate reader (excitation at 545 nm, emission at 590 nm). EC_{50} and CC_{50} values were calculated using GraphPad Prism and are shown as the mean ± SD. Assays were performed in triplicates.

RNA sequencing

Trypanosoma brucei Lister 427 bloodstream forms (0.75×10^6/ml; 15 ml) were treated with $2 \times EC_{50}$ concentrations of either PhenDC3, CX-5461, BMH-21, or no compound (control) for 4 h. Four replicates were prepared for each experimental condition. Total RNA was extracted using the RNeasy Mini Kit (Qiagen, Australia), following the manufacturer's protocol. Briefly, cell pellets were lysed with RNeasy lysis buffer and homogenized using QIAshredder columns. The lysate was subsequently washed with RNeasy wash buffer, and on-column DNase I digestion was performed to remove genomic DNA. RNA was purified and eluted in 50 µl of RNase-free water. RNA quality was assessed using a DeNovix spectrophotometer. Poly(A)-enriched RNA libraries were prepared and sequenced on the Illumina NovaSeq 6000 (S2 200 cycles paired-end reads) at a depth of 15 million reads per sample.

RNA-seq data analysis

The quality of sequence reads was assessed using *FastQC* (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and sequence files were concatenated using the *cat* function (Unix version 8.30). Sequencing adapter sequences were trimmed using *fastp* [63]. Trimmed reads were aligned to the 2018 genome assembly of the Lister 427 strain version 10 (Tb427v10, https://tritrypdb.org) using *Bowtie2* [65]. The alignments were converted from SAM to BAM format, sorted, indexed, and filtered by alignment quality ≥ 1 (-q 1) using *SAMtools* [66]. Reads were quantified using *featureCounts* [79] from the *subread* package (v1.5.2). Differential gene expression analysis and volcano plot visualizations were performed using *edgeR* package in R [80], applying a false discovery rate (FDR) threshold of 0.05 and a $\log_2$ fold change (FC) cut-off of ± 1 (corresponding to a minimum two-fold change in expression). Volcano plots were generated using the *ggplot2* [81] package in R. For the analysis of key genes, gene expression data for each experimental condition were exported as CSV files, and intersected with ChIP-Seq data. GraphPad Prism software was used to visualize the downstream statistical analyses. Multi-dimensional scaling (MDS) plot was generated with *Degust* [82] using min gene read count of 10, FDR threshold of 0.05, and abs log FC cut-off of 1.

Results

G4s are abundant in all trypanosomatids

While current studies often focus on a limited set of representative genomes from trypanosomatids, we have extended the scope of predictive analyses by examining the frequency of putative G-quadruplex sequences (PQSs) per kilobase across the genomes of seven trypanosomatid genera (*Trypanosoma*, *Leishmania*, *Crithidia*, *Phytomonas*, *Herpetomonas*, *Leptomonas*, and *Porcisia*) using the G4Hunter algorithm [55–57]. These genera were selected based on their medical, veterinary, and research relevance, as well as the availability of complete or well-annotated genomes in publicly available databases. This comprehensive approach allows for a broader investigation of G4 distribution within the trypanosomatid family and can be leveraged to highlight changes in their prevalence across different species. For comparative analyses, we also included two *Schistosoma* [53] genomes and the *Homo sapiens* [54] genome as representatives of other parasitic and mammalian organisms. The length of trypanosomatid genomes selected for the analyses varied from ∼17 Mb (*Phytomonas francai*, smallest) to 89 Mb (*T. cruzi*, largest), as shown in Supplementary Table S1. In contrast, the genome of schistosomes and humans is ∼4–35 times larger (400 Mb and 3.1 Gb, respectively). As displayed in Fig. 1A, we observed high variability in PQS prevalence across genera, with *Herpetomonas* and *Leishmania* species exhibiting the highest G4 frequencies and *Schistosoma* showing comparatively the lowest G4 prevalence. Specifically, the highest frequency of G4s (as number of PQSs per kbp) was found in *Herpetomonas tarakana*, which exhibited a 27-fold higher abundance of G4s (frequency of 5.63) compared with *Schistosoma japonicum* (frequency of 0.22) and a six-fold higher abundance than in humans (frequency of 0.93, Fig. 1A). The PQS abundance observed in trypanosomatids compared with schistosomes and humans could, in part, be explained by their higher GC content (average GC% = 52.9%) relative to *Schistosoma* (average GC% = 32.9%) and humans (GC% = 37.8), and therefore a higher propensity to form G4 structures. Visualizing the GC content against the frequency of PQSs revealed distinct patterns across genera (Fig. 1B), showing a near-linear relationship between GC richness of a given genome and its G4 prevalence (R-squared value of 0.709). Interestingly, *Leishmania* species exhibited little variation in G4 prevalence across all evaluated species (Fig. 1B), suggesting potential evolutionary conservation of G4s within this genus that may reflect their relevance to the biology of the parasite. In contrast, *Trypanosoma* species showed greater variability in G4 frequencies (inset zoom in Fig. 1B), suggesting potential differences in the biological and therapeutic significance of G4s across these species. The variation in G4 density observed within the *Trypanosoma* genus suggests that G4 prevalence does not always

Downloaded from https://academic.oup.com/nar/article/54/12/gkag637/8721998 by University of York user on 17 July 2026

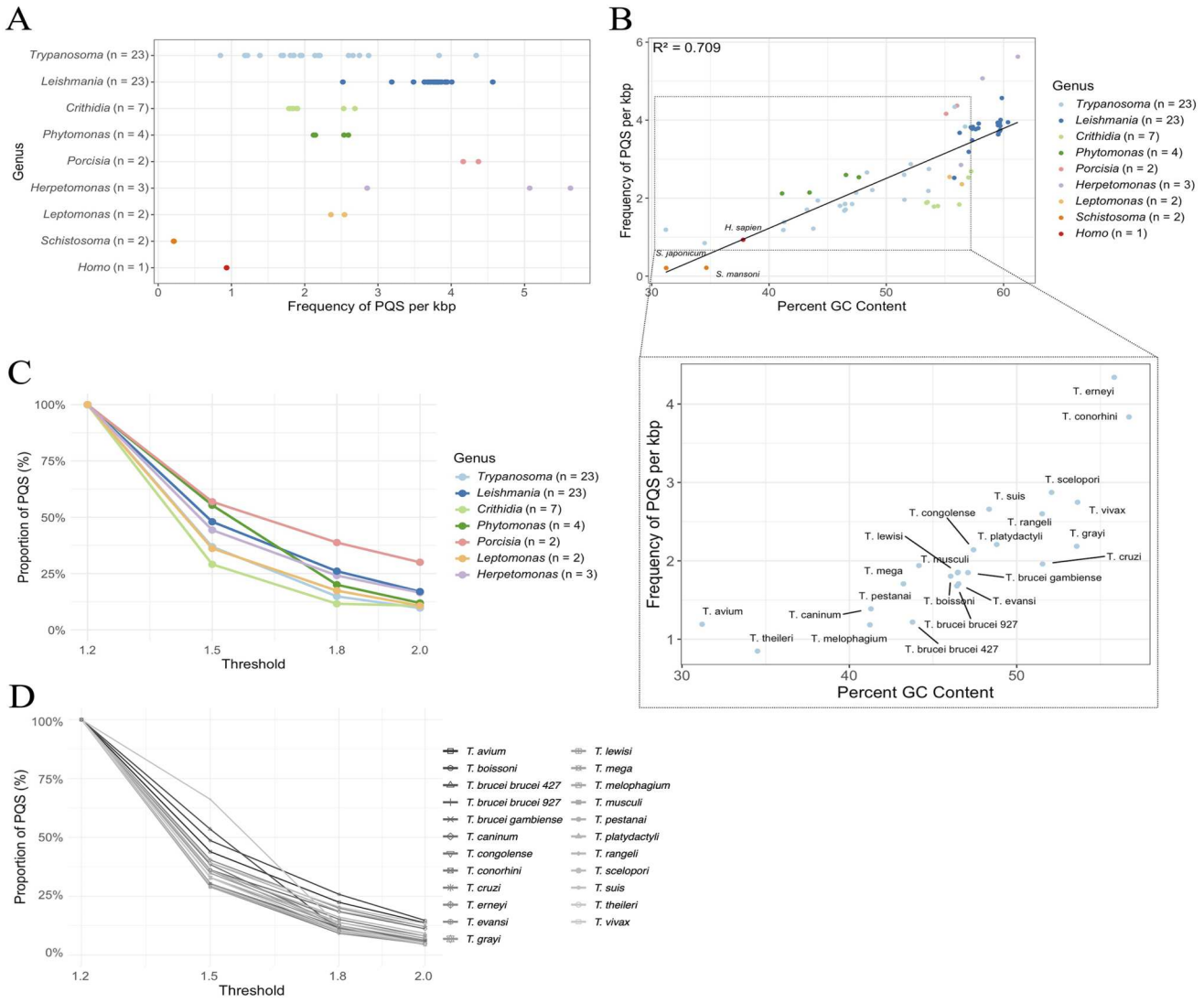


Figure 1. Whole-genome *in silico* analysis of putative G4-forming structures across trypanosomatid species. (**A**) Frequency of PQSs per kilobase (kb). Each dot represents individual species genomes within the analysed trypanosomatid genera, as specified on the *y*-axis. The number of species analysed per genus is shown in parentheses. For comparison, *Schistosoma* and *H. sapiens* genomes are included as reference controls. (**B**) Frequency of PQSs plotted against GC content of each trypanosomatid species analysed. The main graph shows results for all species, while the inset zooms in on *Trypanosoma* species, highlighting the variability in PQS frequency and GC content. (**C**) Change in the proportion of PQSs retained at increasing G4Hunter thresholds (1.2, 1.5, 1.8, and 2.0), expressed as a percentage (%) relative to the mean number of PQSs identified at the lowest threshold (1.2). For each trypanosomatid genus, the mean PQS count at each threshold was divided by the mean count at threshold 1.2 and multiplied by 100, highlighting the sensitivity of PQS detection to threshold stringency. (**D**) Change in the proportion of PQSs retained at increasing G4Hunter thresholds (1.2, 1.5, 1.8, and 2.0) for the 23 *Trypanosoma* species evaluated.

correlate with the GC richness of these parasites, supporting the notion of a potential need of these structures for the biology of certain species. For example, *T. cruzi* and *T. rangeli* have comparable GC content (51.55% and 51.52%, respectively), whilst displaying a 1.3-fold difference in G4 prevalence (i.e. 1.96 and 2.60 PQS frequency per kbp, respectively). The variability in G4 distribution detected suggests that the GC content alone cannot be used to reliably predict the abundance of G4s in each species, which might be determined by other factors. For example, while both species can infect humans, *T. rangeli* is non-pathogenic to mammals [83], suggesting that variations in G4 abundance may be linked to differences in their biological functions, potentially regulating pathogenic potential or disease-causing mechanisms.

To assess the propensity of DNA sequences to form G4 structures, we performed additional analyses by adjusting the threshold values in G4Hunter, which quantifies the likelihood of a DNA sequence to form a G4 structure based on its guanine content and arrangement [55]. Higher G4Hunter threshold values correspond to sequences with a stronger likelihood of forming stable G4 structures; conversely, a lower threshold value predicts G4 structures that are, in general, less stable [53, 55]. As shown in Fig. 1C, increasing G4Hunter thresholds from 1.2 to 2.0 resulted in a progressive decrease in the sensitivity of the analysis (i.e. proportion of PQS found). *Porcisia* retains the largest proportion of PQSs across all thresholds, indicating a high genomic density of sequences capable of forming stable G4s compared with the other genera

Downloaded from https://academic.oup.com/nar/article/54/12/gkag637/8721998 by University of York user on 17 July 2026

analysed. It is worth noting that *Porcisia* does not have the highest GC content among the species analysed, suggesting that GC richness alone is not the primary determinant of G4 abundance. This raises the possibility that additional factors, such as local sequence organization or lineage-specific evolutionary pressures, may contribute to G4 prevalence in these parasites. Across the 23 *Trypanosoma* species analysed, the overall PQS retention pattern was highly similar, with a consistent reduction in retained sequences as G4Hunter stringency increased (Fig. 1D). No individual species displayed a marked enrichment of high-scoring PQSs at stringent thresholds (≥1.8–2.0), suggesting that the distribution of strong G4-forming sequences is broadly comparable across the genus.

T. brucei G4s are enriched in coding regions and TSS

Given the similar PQS abundance in human and *Trypanosoma* cells, and the variability of G4 prevalence among the 23 *Trypanosoma* species examined—despite their comparable GC content—we decided to experimentally map the genomic distribution of G4s in *T. brucei* to gain further insight into their biological roles and potential contribution to pathogenicity. To achieve this, we adapted an established G4 ChIP-Seq protocol [60] to enable the first detection of G4s within the chromatin of bloodstream forms of *T. brucei* parasites. Briefly, chromatin was extracted from *Trypanosoma* cells following an established protocol [61, 62] and crosslinked with formaldehyde to preserve its native architecture. The concentration of recombinant G4-selective antibody, BG4 [59], was optimized empirically to enhance the enrichment of G4-containing chromatin fragments while minimizing non-specific binding. Key refinements to the G4 ChIP-Seq protocol established for human chromatin [60] included adjustments to chromatin-to-antibody ratios (∼5:1), chromatin pre-treatment with Triton X-100 (2% for 10 min), and optimization of sonication cycles (i.e. 20 cycles at 30 s on and 30 s off) to achieve optimal chromatin fragmentation of 100–500 bp. Fine-tuning of these parameters was essential to successfully perform G4 ChIP-Seq on nanograms of trypanosome chromatin, which is imposed by the low *T. brucei* culture yields. Next, we assessed G4 ChIP quality by measuring the relative enrichment of predicted G4-positive *versus* G4-negative regions, which were identified through a pilot ChIP-Seq study (see the 'Materials and methods' section). These regions were located within the bloodstream expression site 1 (BES1), a telomeric locus in *T. brucei* that controls the expression of the active variant surface glycoprotein (VSG) gene, a key factor in immune evasion [64, 84]. Using our adapted G4 ChIP protocol for *T. brucei*, we observed approximately six-fold enrichment of G4-containing regions relative to non-G4 controls across two biological samples, consistent with the ≥5-fold enrichment benchmark established for successful G4 immunoprecipitation in human cells (representative qPCR validation of G4 enrichment across three technical ChIP replicates from biological sample 1 is reported in Fig. 2A) [60]. DNA libraries were then prepared and sequenced using Illumina technology at a depth of 15 million reads per sample.

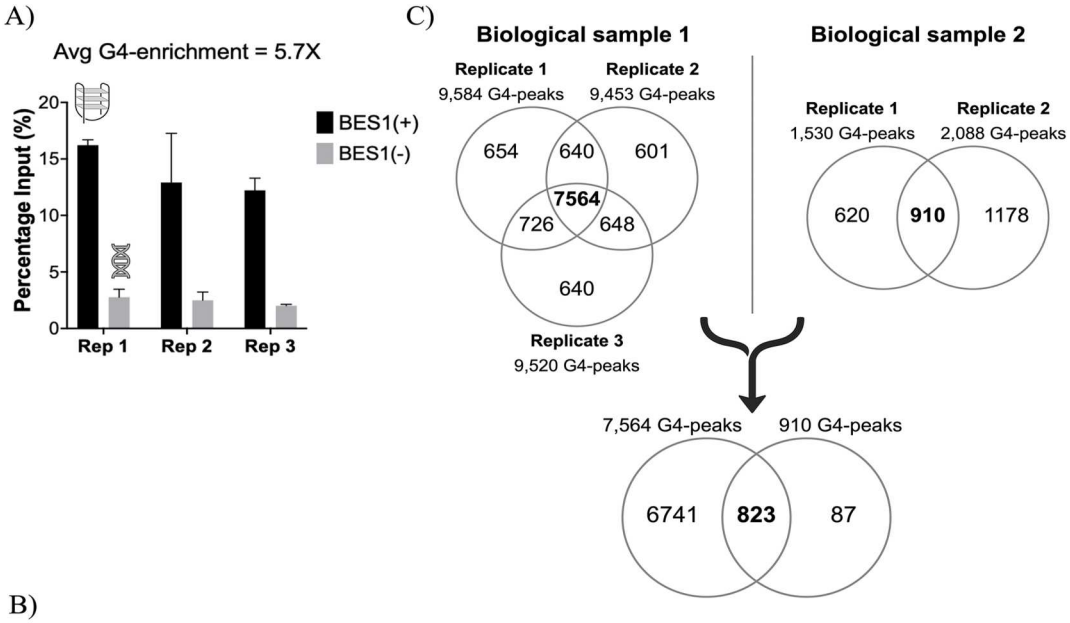
Sequencing data were aligned to the *T. brucei* reference genome (2018 version 10), using *SAMtools* [66] and applying a high mapping quality (MAPQ) threshold of 10 to prioritize uniquely mapped reads. G4-enriched regions were identified with the peak-calling program *MACS2* [67], using parameters [85] optimized for the *T. brucei* genome: size was set to 2.67×10^7, a stringent q-value threshold of ≤ 0.01 (-q 0.01) was used to minimize false positives, and a fixed fragment size of 200 nucleotides (–extsize 200) was applied. This analysis allowed for the detection of G4 peaks (Fig. 2), providing the first comprehensive genomic map of G4s in *T. brucei* chromatin. Our workflow yielded consistent and reproducible G4-enrichment patterns across all ChIP replicates derived from two independent biological samples (Fig. 2B).

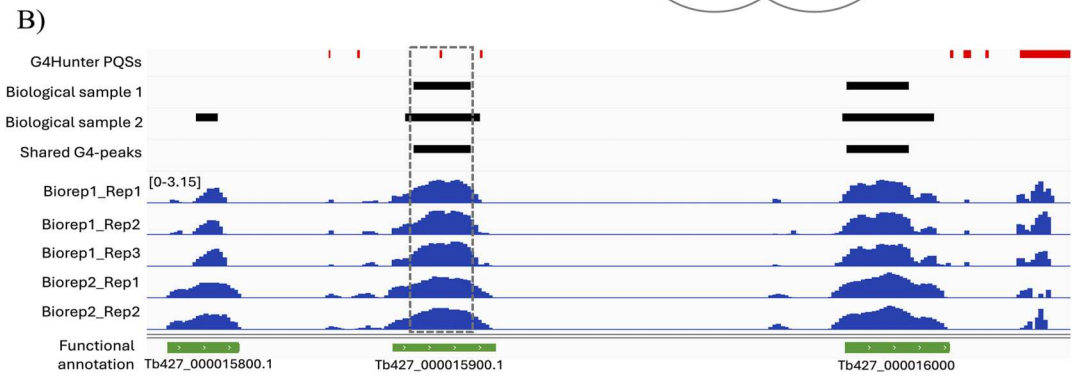
Experimentally derived G4 data were compared with previously published *in silico* predictions and genomic maps of G4s in human and *T. brucei* genomes to assess differences in their distribution patterns [22, 54]. Our G4 ChIP-Seq analysis identified an average of 9519 peaks in *T. brucei* chromatin across three technical replicates for biological sample 1 and an average of 1809 peaks across two technical replicates for biological sample 2 (Fig. 2C). These numbers are substantially lower than the ∼30 000 PQSs predicted by computational analysis [54], consistent with the expectation that only a subset of predicted sequences form stable G4 structures *in vivo* under the experimental conditions. High reproducibility was observed between technical replicates, with 7564 shared peaks (79%) in biological sample 1 and 910 shared peaks (50%) in biological sample 2 (Fig. 2C). When comparing the two independent biological replicates, ∼15% of G4 peaks (i.e. 823 peaks) were shared. Although this overlap is lower than that observed between technical replicates, this is not unexpected given the substantial difference in total peak numbers between the two samples (9519 versus 1809 peaks). Differences in sequencing depth, immunoprecipitation efficiency, or biological variability may contribute to this discrepancy. Importantly, similar variability in peak numbers between biological replicates has also been reported in human G4 ChIP-Seq datasets [86].

We compared G4 peaks identified in biological samples 1 and 2, and the set of peaks shared between the two samples, with PQSs predicted *in silico* using G4Hunter (threshold = 1.2; window size = 25 nucleotides). On average, 33%, 39%, and 45% of the experimentally identified peaks overlapped predicted PQSs in biological sample 1, biological sample 2, and the shared peak set, respectively (Supplementary Table S2). Notably, peaks reproducibly detected in both biological replicates showed the highest overlap with predicted PQSs, supporting the robustness of these experimentally identified G4 sites.

We next compared the genomic distribution of G4s revealed by our ChIP-Seq data to patterns reported previously in studies using *in silico* and *in vitro* approaches. G4-seq performed by Marsico *et al.* [22] on naked single-stranded DNA fragments of the *T. brucei* genome revealed a similar G4 enrichment between the human and the parasite genomes, particularly in promoter and TSS regions. In a subsequent *in silico* study, Puig Lombardi *et al.* [54] intersected the G4 sequences identified by G4-seq [22] with those predicted by G4Hunter and annotated their genomic distribution. This study found a high prevalence of G4s in promoters (42.4%) and exons (31.5%) of the *T. brucei* genome, compared with the high distribution in introns (29.6%) and intergenic regions (21.7%) observed in humans (Fig. 2D) [54]. Although previous studies provided valuable insights into the distribution of G4s in *T. brucei* and humans, direct comparison of genomic locations is challenging due to the unique genomic organization of the parasites. The parasite's genes are organized in long polycistronic transcription units (PTUs), containing dozens

Downloaded from https://academic.oup.com/nar/article/54/12/gkag637/8721998 by University of York user on 17 July 2026





D)

Genomic location	Human* (data from Puig Lombardi et al., *NAR* 2020)	*T. brucei* 927* (data from Puig Lombardi et al., *NAR* 2020)	*T. brucei* 427 G4 ChIP-seq Biological sample 1	*T. brucei* 427 G4 ChIP-seq Biological sample 2	*T. brucei* 427 G4 ChIP-seq Shared peaks	*T. brucei* 927* K^+ (data adapted from Marsico et al., *NAR* 2019)	*T. brucei* 927* K^+/PDS (data adapted from Marsico et al., *NAR* 2019)
3'UTR	1.4	n/a	n/a	n/a	n/a	n/a	n/a
5'UTR	0.5	n/a	n/a	n/a	n/a	n/a	n/a
TTS	2	16.8	4	16	20	8	7
Exon/CDSs	2.1	31.5	88	61	57	20	40
Intron	29.6	n/a	n/a	n/a	n/a	n/a	n/a
Intergenic	21.7	7.5	1	5	3	66	46
Promoter/TSS	4.7	42.4	4	15	16	5	5
Repeats	5.4	n/a	n/a	n/a	n/a	n/a	n/a
LINE/SINE	17.8	n/a	n/a	n/a	n/a	n/a	n/a
Pseudogenes	n/a	n.d.	3	4	4	1	1

Figure 2. G4 ChIP-Seq in trypanosomes. (**A**) Representative G4 enrichment in *T. brucei* chromatin from biological sample 1 assessed by qPCR using primers targeting G4-positive [BES1(+)] and G4-negative [BES1(−)] regions. The average G4 enrichment across three ChIP technical replicates is 5.7-fold relative to control sequences lacking G4 structures. Error bars represent the mean ± SD from two qPCR technical replicates. (**B**) Integrative Genomics Viewer snapshot of the BES1 region (65 589–78 658), showing enrichment of G4-associated peaks identified by ChIP-Seq in biological samples 1 and 2, and peaks shared between these two samples (in black) relative to the reference input control. Peaks were detected using the peak-calling algorithm *MACS2* [67]. Overlapping putative G4-forming sequences (PQSs, in red) were predicted using G4Hunter with a stringency threshold of 1.2 and a 25-nucleotide window. Log_2(ChIP/Input) signal ratios for each ChIP technical replicate, generated using *bamCompare* (bin size of 50 bp) from the *deepTools* [68] suite, are shown in blue. All tracks are displayed on the same scale [0–3.15] to allow direct comparison between samples. The dashed box highlights the G4-positive *locus* targeted by primers for qPCR validation of G4 enrichment post-ChIP Genome functional annotations are shown in green. (**C**) Venn diagram illustrating the number of G4-forming sequences (i.e. G4 peaks) identified in *T. brucei* through G4 ChIP-Seq across two biological samples. Biological sample 1 comprised three technical replicates; biological sample 2 comprised two technical replicates. The diagram shows G4 peaks unique to each biological sample and those shared between them. Technical replicates showed a high degree of G4-peaks overlap (50%–79%), demonstrating high reproducibility of the assay. The two biological samples shared 15% of G4-peaks (i.e. 823 peaks), indicating greater variability between biological samples, consistent with observations in other cell lines [86]. (**D**) Table illustrating the genomic distribution of G4s, as identified through previous experimental and *in silico* analyses in humans and *T. brucei* (data from [54]), and G4-seq approaches (data adapted from [22]) compared with G4 regions identified through ChIP-Seq.

Downloaded from https://academic.oup.com/nar/article/54/12/gkag637/8721998 by University of York user on 17 July 2026

of functionally unrelated genes. Unlike many eukaryotes, *T. brucei* lacks classical promoters for individual genes; instead, transcription initiates at defined TSSs located upstream of these PTUs [87]. Furthermore, genome annotation uses coding DNA sequences (CDSs) rather than exons, as *T. brucei* does not possess the typical exon–intron gene structures common in other eukaryotes. This distinct genomic architecture requires a more rigorous and tailored approach to accurately locate G4s within the parasite's genome based on the ChIP-Seq data. To enhance annotation accuracy, we employed a combined approach whereby G4 ChIP-Seq peaks were initially annotated using the *T. brucei* Lister 427 genome 2018 assembly [64] (version 10, extracted from TriTrypDB) as reference, which includes protein-coding regions, pseudogenes, and non-coding RNAs (ncRNAs). Subsequently, TSSs and TTSs from the most recent genome assembly (v12) [74] were mapped onto the v10 assembly using *Liftoff* [72]. Lastly, peaks that did not overlap any of the previously annotated gene features were classified as intergenic regions. Our analysis revealed a markedly different pattern of G4 distribution in the *T. brucei* genome compared with previous studies. Notably, our data indicate that G4s are predominantly located within CDSs, accounting for 88%, 61%, and 57% of detected peaks in biological samples 1 and 2, and in the shared peak set, respectively (Fig. 2D). In contrast, a smaller fraction was associated with TSS and TTS sites, representing between 4% and 20% of peaks across the three biological datasets (Fig. 2D). The proportions of G4s detected at TSSs differ from those reported by Puig Lombardi *et al*. in *T. brucei*, where 42.4% of G4s (i.e. highest percentage) were assigned to these regions [54], compared with 16% in our shared peak set (Fig. 2D). Additionally, to enable a more direct comparison between our dataset and published G4-seq data under K^+ and K^+/PDS G4-stabilizing conditions from Marsico *et al*. [22], we re-annotated the G4-seq dataset using our tailored approach for genomic annotation described earlier. The G4-seq reannotation revealed a different G4 distribution, with the majority of the G4 peaks being assigned to intergenic regions (i.e. 66% and 46% under K^+ and K^+/PDS G4-stabilizing conditions, respectively) and CDSs (i.e. 20% and 40% under K^+ and K^+/PDS G4-stabilizing conditions, respectively), with only 5%–8% present in TSS/TTS (Fig. 2D). Overall, these data suggest that coding CDSs consistently show high proportion of G4 peaks across all biological datasets analysed, whereas other genomic regions, such as TSSs, TTSs, and intergenic regions, show greater variability.

To further investigate the enrichment pattern of G4 structures in the *T. brucei* genome, we quantified ChIP-Seq peak enrichment across annotated genomic features using a randomization approach [22]. Specifically, we used *bedtools shuffle* to generate three independent sets of randomized peaks across the genome, aiming to preserve the number and length of original peaks but randomizing their locations. By comparing the number of ChIP-Seq G4 peaks to random occurrence, we calculated fold-enrichment values for each genomic feature. The results of this analysis are presented in Table 2. Our analysis confirmed a consistent enrichment within coding regions across all ChIP-Seq samples, with the highest enrichment shown by biological sample 1 with approximately a two-fold increase (1.90 ± 0.03). These results highlight a potential functional role for G4 structures in regulating or interacting with protein-coding sequences in the *T. brucei* genome. The next most enriched features in biological sample 1 were TSS and TTS, exhibiting mean fold enrichments of 1.15 ± 0.03 and 1.13 ± 0.04, respectively. In contrast, TSSs and TTSs were found to be the most enriched genomic features in biological sample 2, with mean fold enrichments of 5.14 ± 1.04 and 4.66 ± 0.57, respectively (Table 2). Analysis of the shared peak set between the two biological samples revealed a similar pattern, with G4 peaks showing approximately five-fold enrichment at TSS and TTS regions, whereas CDSs displayed smaller enrichment (~1-fold). Intergenic regions, pseudogenes, and RNA genes showed little to no enrichment across all ChIP-seq samples, suggesting these regions are less associated with detectable G4 formation. Furthermore, reannotation of the G4-seq data [22] revealed a pattern consistent with that observed in biological sample 2 and the shared G4 peak set, with G4s predominantly enriched at TSSs and TTSs (~2-fold and ~3-fold enrichment, respectively) under both G4-stabilizing conditions (Table 2).

These results suggest that biologically functional G4s in *T. brucei* might be primarily located within CDSs, as well as in TSS and TTS, with lower prevalence in intergenic regions and other genomic categories. Consistent with observations in the human genome, our data highlight the importance of experimental validation of G4 distribution through ChIP-Seq combined with comprehensive genomic annotation to accurately define the G4 landscape of pathogenic organisms with complex or repetitive genome architectures, such as *T. brucei*.

T. brucei chromosome cores are abundant in G4 structures

Using our newly generated G4 map, we next aimed to evaluate the potential biological functions of the identified G4s. A radial plot revealed that the megabase chromosome core regions are densely packed with G4 peaks, consistent with G4 maps previously generated using G4-seq [22] (Fig. 3A). Next, the relationship between G4 distribution and chromosome size in *T. brucei* was investigated using the set of 823 shared G4 peaks identified from ChIP-Seq experiments across the two biological replicates. A strong linear relationship was observed between chromosome size and the number of peaks detected ($R^2 = 0.848$, Supplementary Fig. S1A). Considering that core regions of the megabase chromosomes are known to contain a higher number of gene clusters in comparison to subtelomeric regions, these findings suggest G4s may be relevant in *T. brucei* gene regulation. To rule out the possibility that the high frequency of G4s in the chromosome cores was due to differences in chromosome size or GC content, we normalized the number of G4 peaks by chromosome length and GC content. Specifically, chromosome size (as number of bases) and GC content (as a percentage) were extracted from the FASTA file of the *T. brucei* reference genome using the '*awk*' command. To account for chromosome size, G4 abundance was expressed as peak density (number of G4 peaks per megabase, peaks/Mb), calculated by dividing the number of peaks by chromosome length (Mb). Core chromosomes display relatively consistent G4 peak density, whereas subtelomeric and other chromosomal regions exhibit greater variability, with some loci showing elevated densities (Supplementary Fig. S1B). Normalization by GC content further indicates that GC content alone does not account for the observed distribution of G4 peaks across chromosomes, as shown by the limited variation across regions (Supplementary Fig. S1B).

Downloaded from https://academic.oup.com/nar/article/54/12/gkag637/8721998 by University of York user on 17 July 2026

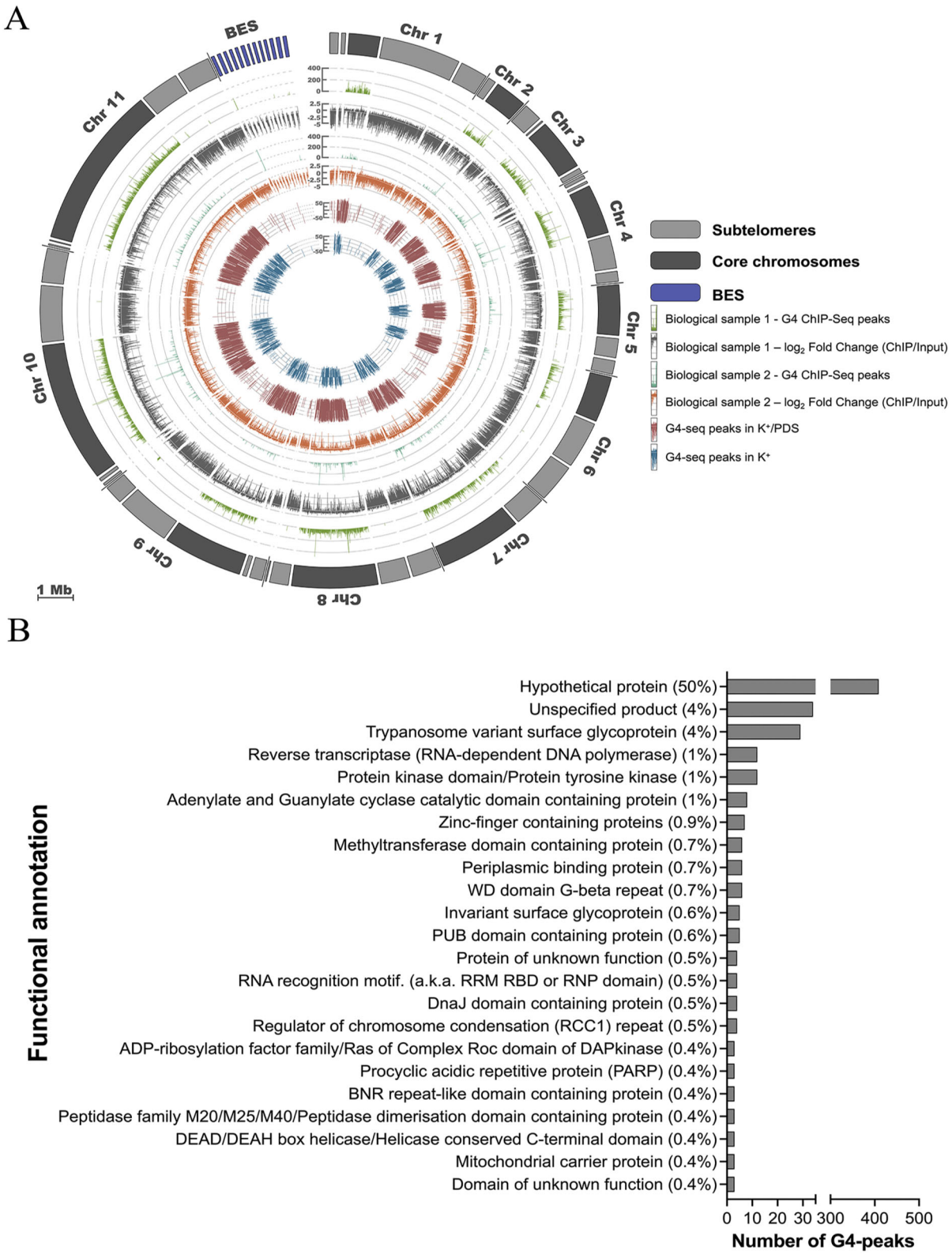


Figure 3. Genomic distribution and functional annotation of G4s in *T. brucei* chromosomes. (**A**) Radial plot of G4-enrichment across the *T. brucei* genome, which consists of 11 megabase chromosomes and bloodstream expression sites (BESs). The outer track (green) displays G4 ChIP-Seq peaks across the genome from biological sample 1, with peak height corresponding to the peak score as determined by *MACS2* analysis. The adjacent inner track (grey) shows the corresponding genome-wide ChIP/Input signal expressed as $\log_2$ fold change, calculated using *BamCompare* with a bin size of 1000 bp. The next inner tracks (turquoise and orange) display the *MACS2* peak score and $\log_2$ fold change ChIP/Input signal for biological sample 2, respectively. The innermost tracks (red and blue) show G4-seq peaks identified under K^+ and K^+/PDS G4-stabilizing conditions, respectively, as reported by Marsico *et al.* [22]. These peaks were remapped from *T. brucei* strain 927 to the 427 v10 genome assembly to enable direct comparison with our dataset. Peak magnitude reflects the mismatch intensity reported in the original study, and strand orientation is indicated by positive and negative values. (**B**) Bar graph showing the most common functional categories associated with G4 peaks identified from ChIP-Seq data. The *x*-axis indicates the number of G4 peaks, and percentages represent the proportion of total peaks assigned to each category.

Downloaded from https://academic.oup.com/nar/article/54/12/gkag637/8721998 by University of York user on 17 July 2026

Table 2. Fold enrichment of G4 peaks detected by ChIP-Seq (biological samples 1 and 2, and their shared peak set) and G4-seq (data from [22]) across different genomic regions, including coding DNA sequences (CDSs), transcriptional start sites (TSSs), transcriptional termination sites (TTSs), intergenic regions, pseudogenes, and non-coding RNAs (ncRNAs)

	G4 ChIP-Seq Biological sample 1			G4 ChIP-Seq Biological sample 2			G4 ChIP-Seq Shared peaks			G4-seq (K^+) Data from Marsico *et al. NAR* 2019			G4-seq (K^+/PDS) Data from Marsico *et al. NAR* 2019		
Features	No. G4-peaks	Mean fold enrichment	SD	No. G4-peaks	Mean fold enrichment	SD	No. of G4-peaks	Mean fold enrichment	SD	No. G4-peaks	Mean fold enrichment	SD	No. G4-peaks	Mean fold enrichment	SD
CDS	6671	1.90	0.03	552	1.35	0.05	467	1.15	0.03	561	0.53	0.02	4090	1.03	0.01
TSS	292	1.15	0.04	132	5.14	1.04	132	4.57	0.70	143	1.88	0.23	558	1.92	0.03
TTS	289	1.14	0.11	149	4.66	0.57	168	5.41	0.91	223	2.91	0.26	747	2.63	0.10
Intergenic	111	0.05	0.00	42	0.14	0.00	24	0.13	0.00	1832	1.52	0.05	4693	1.10	0.02
Pseudogene	198	0.15	0.00	34	0.23	0.02	31	0.19	0.01	35	0.10	0.00	137	0.10	0.00
ncRNA gene	3	0.22	0.08	1	0.83	0.29	1	0.34	0.15	1	0.14	0.03	7	0.38	0.05

The number of peaks per region, the mean fold enrichment, and standard deviation (SD) values are reported.

To infer functional roles of G4s in *T. brucei*, we analysed the most common gene functions associated with the 823 shared G4 peaks identified from ChIP-Seq data (Fig. 3B). Functional classifications were assigned based on current annotations for 2018 genome assembly of the Lister 427 strain (Tb427v10) [64] extracted from TriTrypDB. The largest proportion of these G4 peaks (50%) mapped to genes annotated as hypothetical proteins, indicating that a substantial portion of the G4-enriched regions in *T. brucei* are associated with genes of unknown function. The next most frequent annotations corresponded to unspecified products (4%) and trypanosome VSGs (4%). The presence of G4 peaks within VSG-associated genes is notable, as these proteins play a central role in antigenic variation, allowing trypanosomes to evade the host immune response. Although the functional relevance of these G4s remains to be determined, their association with VSG *loci* raises the possibility that DNA secondary structures may contribute to regulatory processes linked to antigenic variation. This observation is of particular interest, as it could rationalize the reported high potency of G4 ligands in treating trypanosomiasis [43–52]. Additional G4 peaks were associated with RNA-dependent DNA polymerases (reverse transcriptases) (1%), DnaJ domain-containing proteins (0.5%), and DEAD/DEAH box helicases (0.4%), suggesting potential roles for trypanosomal G4s in gene expression regulation and RNA metabolism. G4 peaks were also identified in genes encoding protein kinases (1%), adenylate and guanylate cyclase catalytic domains (1%), and ADP-ribosylation factors (0.4%), suggesting a potential role in cell signalling pathways regulating parasite differentiation and environmental responses. Finally, a small fraction of G4 peaks was associated with methyltransferases (0.7%), suggesting a potential link between G4s and epigenetic regulation.

Our analysis highlighted the enrichment of G4 structures in critical pathways that may serve as potential targets for therapeutic intervention. To explore this further, we aimed to determine whether treatment with G4 ligands, some of which have been previously shown to inhibit parasite growth, could affect these pathways and justify the activity of these ligands as antiparasitic agents. Specifically, we sought to assess whether such treatment leads to changes in the expression profile of these genes, offering insights into how G4 stabilization might affect the growth and regulatory mechanisms of *T. brucei*.

G4-ligands affect the viability of trypanosomes

To evaluate the effect of known G4-targeting ligands on *T. brucei* survival, we performed viability assays using two established G4-ligands, PhenDC3 [78] and CX-5461 [32]. CX-5461 was initially developed as an anticancer agent with activity as an RNA polymerase I (RNA Pol I) inhibitor [88]. Subsequent studies revealed its G4-targeting activity [32]. Interestingly, CX-5461 has been previously reported to kill *T. brucei in vitro* and this activity has been linked to RNA Pol I inhibition rather than G4 stabilization [42]. However, its potential as an antiparasitic agent through G4 targeting has not yet been investigated. Thus, we hypothesised that CX-5461 may act both as a G4 binder and as an RNA Pol I inhibitor. In contrast, PhenDC3 provides a more G4-driven mechanistic view on the antiparasitic potential linked to G4 targeting ligands.

We confirmed that *T. brucei* is sensitive to G4-ligands. CX-5461 exhibited the most potent antiparasitic activity with an EC_{50} of 0.15 ± 0.01 μM (Table 3). However, it also exhibited significant cytotoxicity in human HeLa cells, human colon carcinoma COLO 205 (COLO), and African green monkey foetal kidney epithelial (MA-104) cells, with CC_{50} values of <0.20, 0.53, and 0.88 μM, respectively. This resulted in low selectivity index (S.I.) values (<1.3, 3.5, and 5.9, respectively), indicating poor selectivity for *T. brucei* (Table 3). The low selectivity shown by CX-5461 may be due to its dual mechanism of action, whereby it also inhibits RNA Pol I, leading to potential off-target effects and reduced selectivity in human cells. In contrast, PhenDC3 showed potent antiparasitic activity, with an EC_{50} of 0.36 ± 0.06 μM against *T. brucei* and a favourable selectivity profile. Cytotoxicity values were substantially higher in host cells (HeLa CC_{50}: 53 ± 3 μM; COLO CC_{50}: 21 ± 2 μM; MA-104 CC_{50}: 52 ± 10 μM), resulting in selectivity indices ranging from 52 to 150. These results indicated that targeting G4 structures can translate into selective *T. brucei* killing, supporting the potential of PhenDC3 as a promising molecular scaffold for potential drug development.

Next, we aimed to determine whether the high selectivity displayed by PhenDC3 could be ascribed to changes in expression of genes containing G4s, as detected by ChIP-Seq. To investigate this, we performed RNA-seq and evaluated the transcriptional responses elicited by these ligands on the *T. brucei* transcriptome. Total RNA was extracted from *T. brucei*

Downloaded from https://academic.oup.com/nar/article/54/12/gkag637/8721998 by University of York user on 17 July 2026

Table 3. Antiparasitic activity, cytotoxicity, and selectivity index (S.I.) of PhenDC3 (selective G4 ligand), CX-5461 (dual G4 ligand/RNA polymerase I inhibitor), and BMH-21 (RNA polymerase I inhibitor) in HeLa, human colon carcinoma COLO 205 (COLO), and African green monkey foetal kidney epithelial (MA-104) cells

Compound	*T. brucei* EC_{50} (μM)	*T. brucei* EC_{90} (μM)	HeLa CC_{50} (μM)	S.I. HeLa/Tb	COLO CC_{50} (μM)	S.I. COLO/Tb	MA-104 CC_{50} (μM)	S.I. MA-104/Tb
PhenDC3	0.36 ± 0.06	1.04 ± 0.06	53 ± 3	150	21 ± 2	58	52 ± 10	144
CX-5461	0.15 ± 0.01	0.19 ± 0.01	<0.20*	<1.3	0.53 ± 0.11	3.5	0.88 ± 0.10	5.9
BMH-21	0.11 ± 0.01	0.17 ± 0.06	<0.28*	<2.5	n.d.	-	n.d.	-

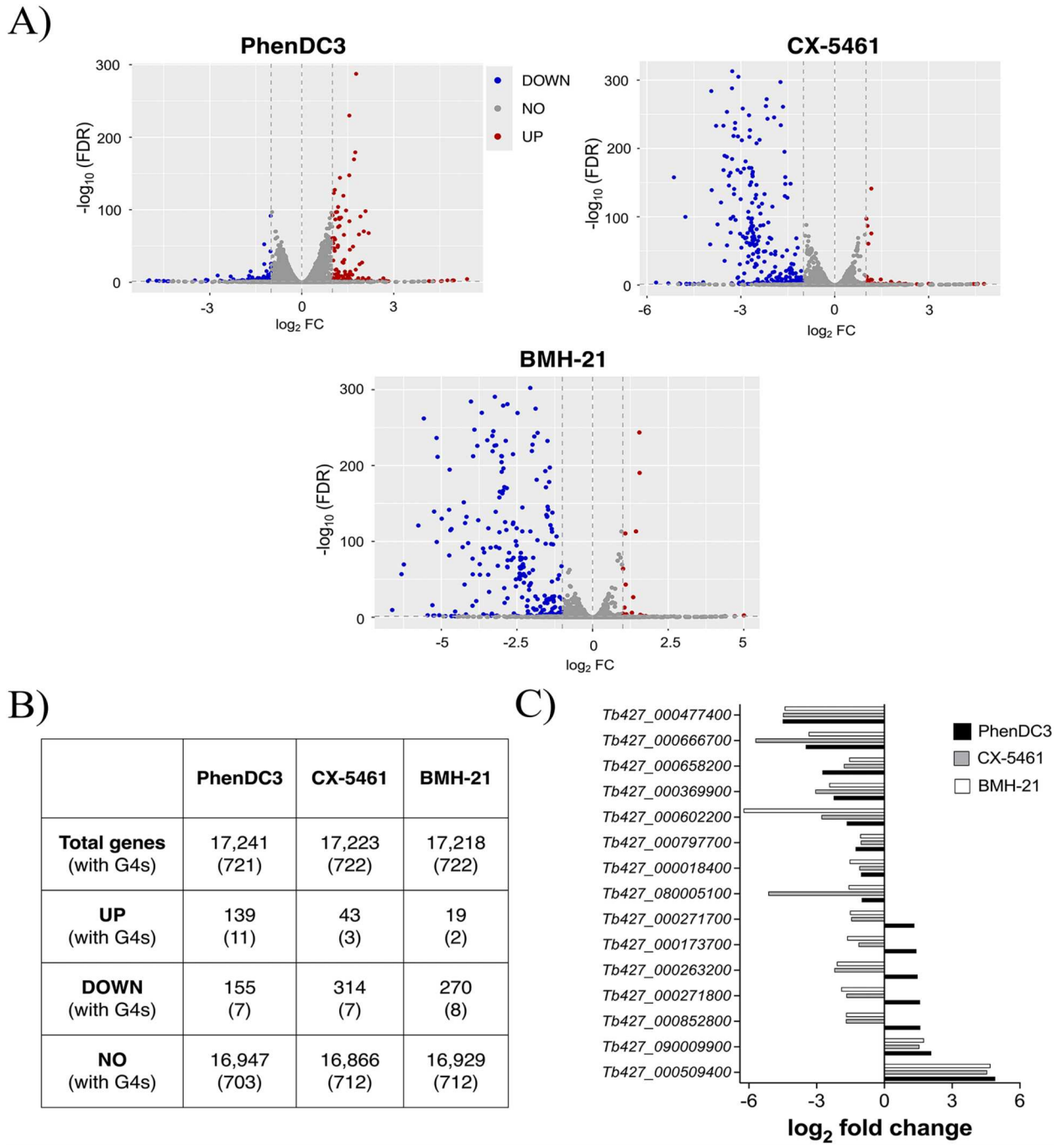
Cell viability of *T. brucei* and HeLa cells was assessed using Alamar Blue (resazurin) as described (Materials and methods). EC_{50} and CC_{50} values were calculated from 11-point concentrations and reported as the mean ± standard deviation (SD). Assays were performed in triplicate. Selectivity Indices (S.I.) were calculated as HeLa, COLO, and MA-104 CC_{50} over *T. brucei* EC_{50} values. *For HeLa cells, lowest dose tested. N.d. = not determined.

cells treated with either CX-5461 or PhenDC3 and compared with an untreated control. To disentangle the transcriptional response mediated by CX-5461 as a G4-binder from RNA Pol I inhibition, a specific RNA Pol I inhibitor, BMH-21, was also included in our analysis [42].

PhenDC3 induces a distinct transcriptome profile compared with CX-5461 and BMH-21

RNA-seq was performed on *T. brucei* bloodstream form parasites after 4 h treatment with PhenDC3, CX-5461, and BMH-21 at concentrations of 2 × EC_{50}. MDS analysis of RNA-seq replicates revealed clear clustering of samples according to treatment condition (control, CX-5461, PhenDC3, and BMH-21), indicating high reproducibility, data quality and distinct transcriptional perturbation elicited by each compound (Supplementary Fig. S2), as shown by samples forming well-separated clusters corresponding to their respective groups. Interestingly, a clear separation for the PhenDC3-treated samples from both the control and the other two treatments was observed, which indicates a difference in gene expression profile. This observation was confirmed by differential gene expression (DGE) analysis, which revealed a distinct gene expression pattern for PhenDC3 compared with CX-5461 and BMH-21, as shown by the volcano plots (Fig. 4A). DGE analysis was performed using a stringent false discovery rate (FDR) threshold of 0.05 and a $\log_2$ fold change (FC) of ± 1 (equivalent to a two-fold change). The FC threshold allowed the identification of a broad set of potentially relevant genes for downstream analysis, while the FDR cut-off of 0.05 provides rigorous control for multiple testing, minimizing the likelihood of false positives. Figure 4B and C summarize the findings from the DGE analysis, which included all gene types (i.e. protein-coding genes, pseudogenes, and ncRNAs). Overall, irrespective of the treatment, ∼98% (∼16 900 genes) of the total genes were not found to be differentially expressed compared with the untreated control. Among the differentially expressed genes (DEGs), CX-5461 appeared to impact the highest number of genes (357), followed by PhenDC3 (294) and BMH-21 (289). Notably, PhenDC3 induced a nearly equal number of upregulated (139) and downregulated (155) genes, while CX-5461 and BMH-21 had a higher prevalence of downregulated genes, an effect that may be due to their inhibitory activity on the RNA Pol I function. Given the G4-targeting potential of PhenDC3, we investigated whether the involvement of G4 structures could explain the different expression profiles observed for this compound. To assess this, we intersected RNA-seq data with the G4 ChIP-Seq dataset comprising the G4 peaks shared between the two biological samples. As shown in Fig. 4B, PhenDC3 resulted in a higher overall number of G4-containing genes being differentially expressed (18 genes total, of which 11 were upregulated and 7 downregulated), compared with BMH-21 (10 total) or CX-5461 (10 total). Furthermore, PhenDC3 induced the strongest gene upregulation overall, including among G4-containing genes (11 genes), suggesting a potential role for G4-associated regulation of gene expression. CX-5461 and BMH-21 triggered broader gene downregulation, with a low proportion of G4-containing genes among downregulated genes, suggesting lower selectivity in targeting of G4s. Furthermore, none of the fifteen genes impacted by all three compounds harboured G4s (Fig. 4C). We also investigated the genes differentially expressed in response to both PhenDC3 and CX-5461, given their reported ability to interact with G4s. Among the 34 genes affected by both compounds (a detailed list of genes with corresponding differential expression is provided in Supplementary Table S3), only one gene (*Tb427_110072300*) contained a G4 motif and was upregulated in both treatments. This overlap, despite being limited to one gene, might reflect that CX-5461 can also act as a G4-ligand for certain targets. However, its global gene expression profile resembles more closely BMH-21, likely due to their shared mechanism of RNA Pol I inhibition. To further disentangle G4-dependent from RNA Pol I-dependent effects, we analysed genes differentially expressed by both CX-5461 and BMH-21, but not by PhenDC3 (a detailed list of genes with corresponding differential expression is provided in Supplementary Table S4). These 196 genes were predominantly downregulated (187) by both compounds and largely lacked G4 peaks (i.e. G4 peaks were found in only six genes), supporting the contribution of RNA Pol I inhibition to the overlapping transcriptional responses observed with CX-5461 and BMH-21. Given the distinct gene upregulation of G4-containing genes observed with PhenDC3, we examined the functional annotation of the 11 upregulated genes harbouring G4 motifs. While a large proportion of *T. brucei* genes are annotated as hypothetical proteins due to the lack of identifiable homologs in sequence databases, the methyltransferase domain and deoxyhypusine synthase annotations were unique to two G4-containing genes upregulated by PhenDC3. In contrast, these domains were not found in any downregulated genes, nor in those upregulated or downregulated by CX-5461 or BMH-21. The presence of a methyltransferase domain in PhenDC3-responsive G4-containing genes is of particular interest, as G4 structures have previously been implicated in the regulation of methylation-related pathways in other eukaryotic organisms [89–91]. Consistent with this, methyltransferases were among the functional categories associated with G4 peaks identified by our ChIP-Seq data, supporting a potential regulatory mechanism involving G4-mediated epi-

Downloaded from https://academic.oup.com/nar/article/54/12/gkag637/8721998 by University of York user on 17 July 2026



	PhenDC3	CX-5461	BMH-21
Total genes (with G4s)	17,241 (721)	17,223 (722)	17,218 (722)
UP (with G4s)	139 (11)	43 (3)	19 (2)
DOWN (with G4s)	155 (7)	314 (7)	270 (8)
NO (with G4s)	16,947 (703)	16,866 (712)	16,929 (712)

Figure 4. PhenDC3 induces a distinctive transcriptome profile in *T. brucei*. (**A**) Volcano plots depicting differential gene expression profiles for PhenDC3, CX-5461, and BMH-21 treatments relative to untreated control (FDR < 0.05 and $\log_2 FC \geq 1$). (**B**) Table summarizing the total number of genes (including protein-coding genes, pseudogenes, and ncRNAs) and the count of DEGs for each treatment condition using an FDR threshold of 0.05 and a $\log_2$ fold change cut-off of $\pm$ 1. UP = upregulated; DOWN = downregulated; NO = no significant change compared with untreated control. Number of genes carrying G4s is indicated in parentheses. (**C**) Bar plot with the fifteen shared genes upregulated or downregulated by PhenDC3, CX-5461, and BMH-21. Their expression levels are indicated as $\log_2$ fold change compared with untreated control.

genetic modulation that may be perturbed upon PhenDC3 treatment.

G4-feature analysis revealed upregulation of G4-containing coding sequences to PhenDC3

Our DGE analysis highlighted a distinct gene expression pattern induced by the G4 ligand PhenDC3 compared with the RNA Pol I inhibitors CX5461 and BMH-21. Furthermore, a higher association between transcriptional dysregulation of genes harbouring G4s was observed for PhenDC3, suggesting a more specific G4-targeting mechanism.

To further evaluate the impact of each treatment on G4-associated genes, we specifically examined differences in gene expression induced by the three compounds on genes identified as carrying G4 motifs based on G4 ChIP-Seq data (us-

Downloaded from https://academic.oup.com/nar/article/54/12/gkag637/8721998 by University of York user on 17 July 2026

ing the set of 823 shared peaks between biological samples 1 and 2). Genes containing G4s were therefore analysed according to their genomic location, grouping them into CDSs, TSSs, TTSs, intergenic, and pseudogenes. As shown in Fig. 5A, the largest differences in gene expression between treatments were observed for genes associated with G4 motifs located in CDSs and TSSs, whereas smaller differences were observed for genes associated with TTSs, intergenic regions, or pseudogenes (Fig. 5A and Supplementary Fig. S3). Treatment with PhenDC3 resulted in a clear upregulation of genes associated with G4s in TSS regions, while more moderate changes were observed for CDS-associated genes. In contrast, genes associated with G4s located at TTSs showed a reduction in expression following PhenDC3 treatment, a trend similar to that observed with BMH-21.

These observations suggest that PhenDC3-induced stabilization of G4 structures near TSS regions may influence transcription initiation or early transcription dynamics, potentially by affecting RNA polymerase recruitment or pausing, thereby increasing gene expression levels. By contrast, at TTS regions—where G4 motifs were found to be enriched in our ChIP-Seq analysis—stabilization of G4 structures by PhenDC3 could potentially enhance transcription termination efficiency or reduce transcriptional readthrough, which would explain the decrease in expression observed for TTS-associated genes.

Together, these observations suggest that the functional impact of G4 stabilization in *T. brucei* may depend on genomic context, with potentially distinct effects at transcription initiation and termination regions. In *T. brucei*, where transcription initiation occurs at strand-switch regions rather than at canonical promoters, such effects may be particularly relevant at TSS clusters. The strong transcriptional response observed for TSS-associated G4-containing genes following PhenDC3 treatment raises the possibility that G4 stabilization may influence transcription initiation sites or early elongation dynamics within polycistronic transcription units. However, the precise mechanisms underlying these effects remain unclear and will require further experimental validation.

To further investigate this hypothesis, we assessed the influence on global gene expression induced by each compound by stratifying genes that do and do not contain G4s. Specifically, we grouped the genes into two main categories: genes that do not contain G4s (*without G4s*) and genes that contain or have proximal G4s (*with G4s*). Given the substantial imbalance in the number of genes between the two categories, genes '*without G4s*' were randomly subsampled using the *awk* command to match the number of genes '*with G4s*,' enabling a more balanced comparison between datasets. Genes '*with G4s*' were then further categorized according to genomic location [i.e. protein coding genes (PCG), TSS, TTS, intergenic, and pseudogenes]. Figure 5B shows fold changes in gene expression relative to the untreated control ($\log_2$ treated/untreated) for genes with and without G4s following exposure to each treatment. Statistical comparisons were performed between gene categories within each treatment condition to determine whether G4-associated genes exhibit differential transcriptional responses compared with genes lacking G4s. Interestingly, we observed that PhenDC3 exhibits a significant difference in expression between genes with and without G4s, especially in coding sequences (PCG) and TSSs (Fig. 5B). In contrast, CX-5461 and BMH-21 show little or no significant difference in expression between genes with and without G4s (Fig. 5B). Interestingly, while CX-5461 and BMH-21 induced an overall downregulation of genes (both without and with G4s), PhenDC3 induced an opposite effect, downregulating genes without G4s and upregulating most gene categories carrying G4s (with the only exception of TTSs). This evidence further supports our hypothesis that treatment with G4-ligands, such as PhenDC3, may act through a G4-mediated mechanism that alters preferentially genes carrying G4s, particularly in PCG and TSS.

Discussion

Our study presents a comprehensive analysis of G-quadruplex (G4) DNA structures in trypanosomatid parasites, providing insights into G4's potential as molecular targets to treat infections caused by these parasites.

Through whole-genome *in silico* analyses, we identified putative G4 sequences (PQSs) across the genomes of 64 trypanosomatid species, revealing intriguing patterns of G4 richness that may reflect different biological functions mediated by G4s in different trypanosomatid species (Fig. 1). For instance, the 23 *Leishmania* species analysed exhibited a consistently high PQS frequency (with an average of 3.74) and a narrow frequency range of 2.52–4.57, suggesting that G4s might have conserved functional roles. Conversely, the 23 *Trypanosoma* species examined showed a broader range of PQS frequencies, ranging from 0.85 to 4.34 PQS/Kb. We observed a near-linear relationship between GC content and PQS frequency across genera, supporting the well-established notion that G4s are more likely to form in GC-rich regions [22]. However, we identified a notable exception to this trend, which suggested that GC content alone does not fully explain the prevalence of predicted G4 structures. *Porcisia* species retained the largest proportion of PQSs at increasing G4Hunter thresholds, despite having a lower GC content compared with other related species, such as *Leishmania* (Fig. 1C). This suggests that other factors affect G4 prevalence in these organisms that go beyond GC richness. For example, *Leishmania* is a medically significant genus that infects both humans and animals, relying on a broad repertoire of cell surface proteins to facilitate the intracellular propagation of amastigotes within vertebrate macrophages—a critical adaptation for their survival and virulence [92]. In contrast, *Porcisia* species, which infect porcupines, have a substantially reduced repertoire of such proteins and do not rely on macrophage infection for their lifecycle [92]. The retention of PQSs in *Porcisia* despite lower GC content might reflect alternative mechanisms regulated by G4s that are tailored around the distinct ecological niche of these species.

To gain further insights on the potential physiological relevance of G4s in these organisms, we next applied G4 ChIP-Seq to map G4s in *T. brucei* chromatin, generating the first chromatin-based map of G4s in this parasite. G4 ChIP identified an average of 5664 G4 peaks consistent across two independent biological replicates, with 9519 peaks detected in replicate 1 and 1809 peaks in replicate 2, confirming the particularly dense distribution of G4 structures in *T. brucei* already predicted computationally. This is particularly relevant when considering the relatively small size of the *T. brucei* genome (26 Mb) compared with other species, such as humans (3.1 Gb). This further highlighted the potential for functional relevance of G4s in *T. brucei* and agrees with the established potential of G4 ligands in acting as antiparasitic agents.

Downloaded from https://academic.oup.com/nar/article/54/12/gkag637/8721998 by University of York user on 17 July 2026

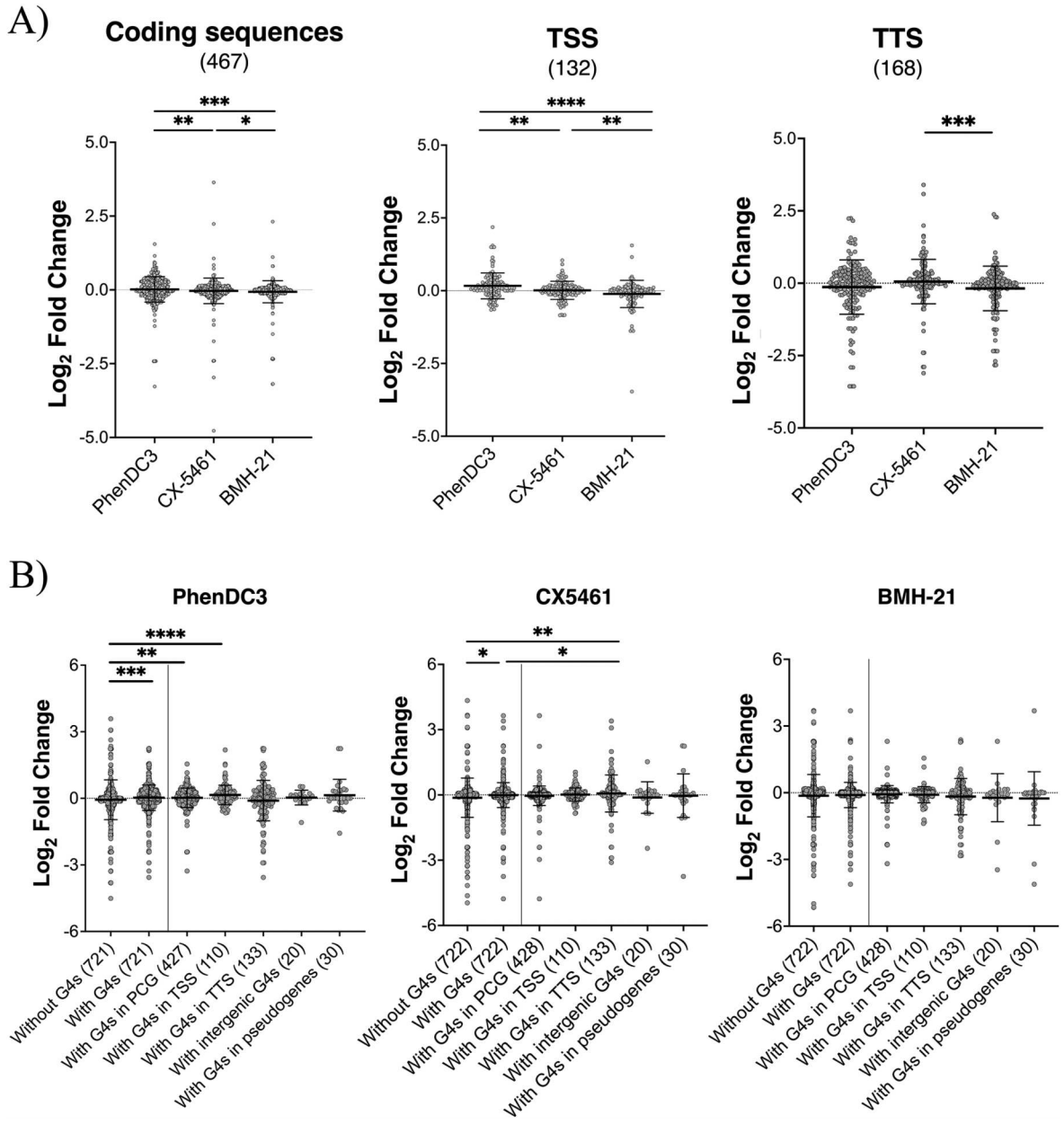


Figure 5. Expression changes of G4-associated genes across compound treatments. (**A**) Gene expression changes of G4-containing genes, as identified by G4 ChIP-Seq, upon treatment with PhenDC3, CX-5461, or BMH-21 in coding sequences, transcriptional start sites (TSS), and transcriptional termination sites (TTS). The number of G4-associated genes per category is shown in parentheses. (**B**) Global changes in gene expression of *T. brucei* genes with and without G4s upon compound treatment. Change in gene expression (as $\log_2$ fold change) of genes in coding sequences (protein-coding genes, PCG), TSS, TTS, intergenic regions, and pseudogenes. The number of genes per category is shown in parentheses. A subset of genes from the 'without G4s' group was subsampled using the *awk* command to match the number of genes in the 'with G4s' group. For both panels (A) and (B), differential gene expression is shown as mean $\log_2$ FC relative to untreated control (baseline = 0) ± SD for each treatment. Statistical significance between groups was determined using the Mann–Whitney U-test, where **** denotes $P \leq 0.0001$, *** denotes $P \leq 0.001$, ** denotes $P \leq 0.01$, * denotes $P \leq 0.05$. Where no symbols are shown, differences are not statistically significant (i.e. $P > 0.05$).

The genome-wide distribution of G4s detected by ChIP-Seq differs markedly from what has been observed in humans, whereby G4s are predominantly observed in introns, intergenic, and promoter regions [22, 54]. In contrast to previous findings, our data reveal a consistent enrichment of *T. brucei* G4s within coding sequences across all biological samples, with comparatively lower and more variable abundances at TSS, TTS, and other regions (Fig. 2D and Table 2). This distribution diverges from that reported by Marsico *et al.* using G4-seq, which showed a pattern more closely resembling the G4 landscape in human genomes when measured on naked single-stranded DNA. These differences likely reflect the distinct experimental approaches employed. G4-seq is based on sequencing extracted genomic DNA under conditions that promote G4 formation (e.g. potassium ions and/or G4-stabilizing ligands) and identifies G4s by detecting mismatches between sequencing reads generated under these conditions. Our ChIP-based approach, in contrast, captures G4s within native chromatin extracted directly from *T. brucei* parasites. As such, it may better reflect the physiological and dynamic nature of G4

Downloaded from https://academic.oup.com/nar/article/54/12/gkag637/8721998 by University of York user on 17 July 2026

formation *in vivo*. Interestingly, despite these methodological differences, the total number of G4s identified in our study (5664 peaks) is broadly comparable to the range reported by Marsico *et al.* [22]. However, the genomic distribution of G4s differs markedly between the two datasets, which may also result from differences in annotation strategies. While Marsico *et al.* applied a cross-species framework, our analysis was tailored to the unique genomic architecture of *T. brucei*, including its organization into polycistronic transcription units. Re-annotation of the G4-seq dataset using our combined approach revealed similar enrichment patterns (Table 2).

Altogether, these observations underscore the influence of experimental context—particularly the chromatin state and genome annotation—on G4 detection. G4-seq, which uses naked single-stranded DNA, does not account for chromatin structure or other dynamic factors that may regulate G4 folding and stability in a cellular context. By contrast, ChIP-Seq approaches preferentially capture G4s that are stably folded under physiological conditions, providing a potentially more biologically relevant snapshot of G4 dynamics in living cells.

The high density of G4s in chromosome cores of the *T. brucei* genome suggests that these structures might be involved in the regulation of key biological functions of the parasite, including transcriptional regulation of genes. Therefore, we investigated the effects of established G4-ligands on parasite viability and transcriptional activity. As shown by the data reported in Table 3, PhenDC3 exhibited potent antiparasitic activity with a favourable selectivity profile, making it a strong candidate for further antiparasitic drug development. In contrast, the dual G4-ligand and RNA Pol I inhibitor CX-5461 showed high anti-trypanosomal efficacy but low selectivity, which is in agreement with its use in humans as an anticancer agent. The reduced selectivity observed for CX-5461 likely reflects, at least in part, its inhibitory effect on RNA Pol I activity, an essential process in eukaryotic cells. In support of this interpretation, CX-5461 exhibited transcriptional responses that more closely resembled those induced by the specific RNA Pol I inhibitor BMH-21 than those observed with PhenDC3, indicating that RNA Pol I inhibition substantially contributes to its global effects. These data suggest that *T. brucei* is highly sensitive to treatment with G4 ligands and, given the favourable selectivity profile observed for PhenDC3, support the therapeutic potential of G4-targeting approaches in treating trypanosome infections. This is consistent with recent reports describing the anti-trypanosomal activity of additional G4-ligands [42–52].

The distinct transcriptome profiles following PhenDC3, CX-5461, and BMH-21 exposure revealed by RNA-seq further support their different modes of action. The ability of PhenDC3 to induce changes in the expression of G4-containing genes supports the hypothesis of a relevant role of G4s in regulating gene expression levels in *T. brucei* (Figs 4 and 5), which is in agreement with what has been extensively reported in human cells [93–95]. The significant differential gene expression observed in genes with and without G4s, especially in PCG and TSSs following PhenDC3 treatment, further supports the hypothesis that G4s may act as modulators of gene expression in trypanosomes (Fig. 5B). Targeting G4 structures offers a promising strategy for perturbing gene expression homeostasis in pathogens like *T. brucei*, which can be potentially leveraged to overcome the increasing incidence of resistance against current therapeutic agents. We anticipate that this approach could also have broader applications in combating infectious diseases beyond parasites and might be explored in broader microbiology contexts.

The distinct upregulation of G4-containing genes observed following PhenDC3 treatment is consistent with previous observations in trypanosomes, where drug-induced transcriptional changes frequently include widespread gene upregulation [96, 97]. The PhenDC3 effect might be explained by the different genomic distribution of G4s in *T. brucei* compared with mammals. Additionally, this could also reflect different mechanisms of G4-protein recognition leveraged in *T. brucei* that might be enhanced in the presence of the ligand, leading to gene expression stimulation. This would be in line with our recent finding demonstrating that G4 ligands can act as molecular glues and increase G4-protein interactions at certain G4-sites [98]. From a more functional perspective, the upregulation of G4-containing genes associated with methyltransferase domain-containing proteins, which is unique to PhenDC3 and absent for CX-5461 or BMH-21, warrants further investigation and highlights potential for epigenetic relevance in parasite biology. Although methyltransferase-associated G4 peaks constitute a modest fraction of total G4 ChIP-Seq peaks (10 identified), their presence suggests G4 formation at these *loci*. Furthermore, RNA-seq analysis identified 2 methyltransferase genes, one of which harboured G4 motifs, that were specifically upregulated by PhenDC3 but remained unaffected by the other two compounds. This confirms a link between G4 stabilization and epigenetic regulation mediated by methyltransferase activity, which has been previously described in other organisms [90, 99]. In the context of therapeutic development, gene expression upregulation may reflect broader disturbances of transcriptional homeostasis rather than selective induction of beneficial or detrimental pathways. We therefore interpret the upregulation observed upon Phen-DC3 treatment as evidence of disrupted gene expression in trypanosomes. While such disruption may contribute to parasite killing, the precise roles of specific upregulated genes require further investigation.

Overall, this study provided the first chromatin-based genome-wide map of G4s in *T. brucei*, highlighting species-specific distribution, transcriptional relevance, and therapeutic potential of G4s in trypanosomes. The distinct G4 enrichment pattern observed in *T. brucei*, compared with computational predictions and other mapping methods, reinforces the importance of experimental validation using physiologically relevant approaches when evaluating G4 prevalence—an observation previously made in human cells. Future studies should now focus on exploring the mechanistic roles of unknown proteins that have shown potential for gene expression modulation by means of G4-targeting, as this might lead to the rational design of a new generation of therapeutic agents. Additionally, further optimization of current G4 ligands with improved selectivity profiles holds promise for the development of new targeted therapies against trypanosome infections.

Supplementary data

Supplementary data is available at NAR online.

Acknowledgements

We acknowledge Prof. Gloria Rudenko and her research group at Imperial College London for invaluable guidance through the initial phases of the study.

Downloaded from https://academic.oup.com/nar/article/54/12/gkag637/8721998 by University of York user on 17 July 2026

Author contributions: Ludovica Monti (Conceptualization [supporting], Data curation [lead], Formal analysis [lead], Methodology [lead], Project administration [supporting], Writing – original draft [equal]), Genta Firth (Data curation [supporting]), Joana R. Correia Faria (Methodology [supporting], Resources [supporting], Writing – review & editing [supporting]), John. M Kelly (Data curation [supporting], Resources [supporting]), Gem Flint (Methodology [supporting]), Silvia Galli (Methodology [supporting]), Thomas E. Maher (Resources [supporting]), Htay Mon Aye (Methodology [supporting]), and Marco Di Antonio (Conceptualization [lead], Formal analysis [equal], Funding acquisition [lead], Project administration [lead], Writing – original draft [equal], Writing – review & editing [equal])

Conflict of interest

None declared.

Funding

M.D.A. is supported by a Biotechnology and Biological Sciences Research Council (BBSRC) David Phillips Fellowship (BB/R011605/1) and is a Lister Institute Research Prize holder (2022). L.M. was supported by funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 101027645. J.F.R.C. is supported by a Wellcome Trust & Royal Society Sir Henry Dale Fellowship (222573/Z/21/Z). T.E.M. was supported by the Engineering and Physical Sciences Research Council (EPSRC, EP/S023518/1).

Data availability

ChIP-Seq and RNA-seq data have been deposited at Gene Expression Omnibus (GEO) under accession numbers GSE302611 and GSE302610.

References

1. World Health Organization. Trypanosomiasis, human African (sleeping sickness). 2023. Available at: https://www.who.int/news-room/fact-sheets/detail/trypanosomiasis-human-african-(sleeping-sickness). Date accessed 1 April 2026.
2. Büscher P, Cecchi G, Jamonneau V *et al.* Human African trypanosomiasis. *Lancet* 2017;390:2397–409.
3. Kennedy PG. Human African trypanosomiasis of the CNS: current issues and challenges. *J Clin Invest* 2004;113:496–504. https://doi.org/10.1172/JCI200421052
4. Venturelli A, Tagliazucchi L, Lima C *et al.* Current treatments to control African trypanosomiasis and One Health perspective. *Microorganisms* 2022;10:1298. https://doi.org/10.3390/microorganisms10071298
5. De Rycker M, Wyllie S, Horn D *et al.* Anti-trypanosomatid drug discovery: progress and challenges. *Nat Rev Microbiol* 2023;21:35–50. https://doi.org/10.1038/s41579-022-00777-y
6. Okello I, Mafie E, Eastwood G *et al.* African animal trypanosomiasis: a systematic review on prevalence, risk factors and drug resistance in Sub-Saharan Africa. *J Med Entomol* 2022;59:1099–143. https://doi.org/10.1093/jme/tjac018
7. Kasozi KI, MacLeod ET, Ntulume I *et al.* An update on African trypanocide pharmaceutics and resistance. *Front Vet Sci* 2022;9:828111. https://doi.org/10.3389/fvets.2022.828111
8. Ungogo MA, Koning HP. Drug resistance in animal trypanosomiases: epidemiology, mechanisms and control strategies. *Int J Parasitol Drugs Drug Resist* 2024;25:100533. https://doi.org/10.1016/j.ijpddr.2024.100533
9. Pelfrene E, Harvey Allchurch M, Ntamabyaliro N *et al.* The European Medicines Agency's scientific opinion on oral fexinidazole for human African trypanosomiasis. *PLoS Negl Trop Dis* 2019;13:e0007381. https://doi.org/10.1371/journal.pntd.0007381
10. Lindner AK, Lejon V, Chappuis F *et al.* New WHO guidelines for treatment of gambiense human African trypanosomiasis including fexinidazole: substantial changes for clinical practice. *Lancet Infect Dis* 2020;20:e38–46. https://doi.org/10.1016/S1473-3099(19)30612-7
11. Lindner AK, Lejon V, Barrett MP *et al.* New WHO guidelines for treating rhodesiense human African trypanosomiasis: expanded indications for fexinidazole and pentamidine. *Lancet Infect Dis* 2024;25:e77–85. https://doi.org/10.1016/S1473-3099(24)00581-4
12. Alsan M. 2015; The effect of the tsetse fly on African development. *Am Econ Rev* 2015;105:382–410. https://doi.org/10.1257/aer.20130604
13. Tora E, Dana D. Epidemiology and economic cost of trypanosomosis among smallholder cattle herders in Arba Minch and Zuria Districts. Gamo Zone, Ethiopia. *Environ Health Insights* 2024;18:11786302241274698. https://doi.org/10.1177/11786302241274698
14. Food and Agriculture Organization of the United Nations. Programme Against African Trypanosomosis (PAAT). Available at: https://www.fao.org/paat/ . Date accessed 1 April 2026.
15. Kasozi KI, MacLeod ET, Welburn SC. Systematic review and meta-analysis on human African trypanocide resistance. *Pathogens* 2022;11:1100. https://doi.org/10.3390/pathogens11101100
16. Davis JT. G-quartets 40 years later: from 5′-GMP to molecular biology and supramolecular chemistry. *Angew Chem Int Ed* 2004;43:668–98. https://doi.org/10.1002/anie.200300589
17. Rhodes D, Lipps HJ. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res* 2015;43:8627–37. https://doi.org/10.1093/nar/gkv862
18. Spiegel J, Adhikari S, Balasubramanian S. The structure and function of DNA G-quadruplexes. *Trends Chem* 2020;2:123–36. https://doi.org/10.1016/j.trechm.2019.07.002
19. Huppert JL, Balasubramanian S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res* 2007;35:406–13. https://doi.org/10.1093/nar/gkl1057
20. Chambers VS, Marsico G, Boutell JM *et al.* High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat Biotechnol* 2015;33:877–81. https://doi.org/10.1038/nbt.3295
21. Sahakyan AB, Chambers VS, Marsico G *et al.* Machine learning model for sequence-driven DNA G-quadruplex formation. *Sci Rep* 2017;7:14535. https://doi.org/10.1038/s41598-017-14017-4
22. Marsico G, Chambers VS, Sahakyan AB *et al.* Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res* 2019;47:3862–74. https://doi.org/10.1093/nar/gkz179
23. Marquevielle J, Robert C, Lagrabette O *et al.* Structure of two G-quadruplexes in equilibrium in the KRAS promoter. *Nucleic Acids Res* 2020;48:9336–45. https://doi.org/10.1093/nar/gkaa387
24. Ambrus A, Chen D, Dai J *et al.* Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry* 2005;44:2048–58. https://doi.org/10.1021/bi048242p
25. Phan AT, Kuryavyi V, Burge S *et al.* Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *J Am Chem Soc* 2007;129:4386–92. https://doi.org/10.1021/ja068739h
26. Robinson J, Raguseo F, Nuccio SP *et al.* DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nucleic Acids Res* 2021;49:8419–31. https://doi.org/10.1093/nar/gkab609

Downloaded from https://academic.oup.com/nar/article/54/12/gkag637/8721998 by University of York user on 17 July 2026

27. Hänsel-Hertsch R, Beraldi D, Lensing SV *et al.* G-quadruplex structures mark human regulatory chromatin. *Nat Genet* 2016;48:1267–72. https://doi.org/10.1038/ng.3662
28. Hänsel-Hertsch R, Simeone A, Shea A *et al.* Landscape of G-quadruplex DNA structural regions in breast cancer. *Nat Genet* 2020;52:878–83. https://doi.org/10.1038/s41588-020-0672-8
29. Robinson J, Flint G, Garner I *et al.* G-quadruplex structures regulate long-range transcriptional reprogramming to promote drug resistance in ovarian cancer cells. *Genome Biol* 2025;26:183. https://doi.org/10.1186/s13059-025-03654-y
30. Pirota V, Stasi M, Benassi A *et al. Annual Reports in Medicinal Chemistry*. Vol. 54. Elsevier, 2020, 163–96. https://doi.org/10.1016/bs.armc.2020.04.008
31. Figueiredo J, Mergny JL, Cruz C. G-quadruplex ligands in cancer therapy: progress, challenges, and clinical perspectives. *Life Sci* 2024;340:122481. https://doi.org/10.1016/j.lfs.2024.122481
32. Xu H, Di Antonio M, McKinney S *et al.* CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nat Commun* 2017;8:14432. https://doi.org/10.1038/ncomms14432
33. Kosiol N, Juranek S, Brossart P *et al.* G-quadruplexes: a promising target for cancer therapy. *Mol Cancer* 2021;20:40. https://doi.org/10.1186/s12943-021-01328-4
34. Ahmed AA, Greenhalf W, Palmer DH *et al.* The potent G-quadruplex-binding compound QN-302 downregulates S100P gene expression in cells and in an *in vivo* model of pancreatic cancer. *Molecules* 2023;28:2452. https://doi.org/10.3390/molecules28062452
35. Dumetz F, Merrick CJ. Parasitic protozoa: unusual roles for G-quadruplexes in early-diverging eukaryotes. *Molecules* 2019;24:1339. https://doi.org/10.3390/molecules24071339
36. Cueny RR, McMillan SD, Keck JL. G-quadruplexes in bacteria: insights into the regulatory roles and interacting proteins of non-canonical nucleic acid structures. *Crit Rev Biochem Mol Biol* 2022;57:539–61. https://doi.org/10.1080/10409238.2023.2181310
37. Pathak R. G-quadruplexes in the viral genome: unlocking targets for therapeutic interventions and antiviral strategies. *Viruses* 2023;15:2216. https://doi.org/10.3390/v15112216
38. Ma X, Feng Y, Luo Z *et al.* Identification and functional characterization of G-quadruplexes in plants. *Trends Plant Sci* 2022;27:952–3. https://doi.org/10.1016/j.tplants.2022.03.015
39. Harris LM, Merrick CJ. G-quadruplexes in pathogens: a common route to virulence control? *PLoS Pathog* 2015;11:e1004562. https://doi.org/10.1371/journal.ppat.1004562
40. Harris LM, Monsell KR, Noulin F *et al.* G-quadruplex DNA motifs in the malaria parasite *Plasmodium falciparum* and their potential as novel antimalarial drug targets. *Antimicrob Agents Chemother* 2018;62:10–1128. https://doi.org/10.1128/AAC.01828-17
41. Monti L, Di Antonio M. G-quadruplexes as key transcriptional regulators in neglected trypanosomatid parasites. *ChemBioChem* 2023;24:e202300265. https://doi.org/10.1002/cbic.202300265
42. Kerry LE, Pegg EE, Cameron DP *et al.* Selective inhibition of RNA polymerase I transcription as a potential approach to treat African trypanosomiasis. *PLoS Negl Trop Dis* 2017;11:e0005432. https://doi.org/10.1371/journal.pntd.0005432
43. Belmonte-Reche E, Martínez-García M, Guédin A *et al.* G-quadruplex identification in the genome of protozoan parasites points to naphthalene diimide ligands as new antiparasitic agents. *J Med Chem* 2018;61:1231–40. https://doi.org/10.1021/acs.jmedchem.7b01672
44. Guillon J, Cohen A, Das RN *et al.* Design, synthesis, and antiprotozoal evaluation of new 2,9-bis[(substituted-aminomethyl)phenyl]-1,10-phenanthroline derivatives. *Chem Biol Drug Des* 2018;91:974–95. https://doi.org/10.1111/cbdd.13164
45. Zuffo M, Stucchi A, Campos-Salinas J *et al.* Carbohydrate-naphthalene diimide conjugates as potential antiparasitic drugs: synthesis, evaluation and structure-activity studies. *Eur J Med Chem* 2019;163:54–66. https://doi.org/10.1016/j.ejmech.2018.11.043
46. Guillon J, Cohen A, Boudot C *et al.* Design, synthesis, and antiprotozoal evaluation of new 2,4-bis[(substituted-aminomethyl)phenyl]quinoline, 1,3-bis[(substituted-aminomethyl)phenyl]isoquinoline and 2,4-bis[(substituted-aminomethyl)phenyl]quinazoline derivatives. *J Enzyme Inhib Med Chem* 2020;35:432–59. https://doi.org/10.1080/14756366.2019.1706502
47. Street STG, Peñalver P, O'Hagan MP *et al.* Imide condensation as a strategy for the synthesis of core-diversified G-quadruplex ligands with anticancer and antiparasitic activity. *Chemistry A European J* 2021;27:7712–21. https://doi.org/10.1002/chem.202100040
48. Guillon J, Cohen A, Boudot C *et al.* Design, synthesis, and antiprotozoal evaluation of new promising 2, 9-bis (substituted-aminomethyl)-4, 7-phenyl-1, 10-phenanthroline derivatives, a potential alternative scaffold to drug efflux. *Pathogens* 2022;11:1339. https://doi.org/10.3390/pathogens11111339
49. Belmonte-Reche E, Benassi A, Peñalver P *et al.* Thiosugar naphthalene diimide conjugates: G-quadruplex ligands with antiparasitic and anticancer activity. *Eur J Med Chem* 2022;232:114183. https://doi.org/10.1016/j.ejmech.2022.114183
50. Pérez-Soto M, Peñalver P, Street STG *et al.* Structure-activity relationship studies on divalent naphthalene diimide G quadruplex ligands with anticancer and antiparasitic activity. *Bioorg Med Chem* 2022;71:116946. https://doi.org/10.1016/j.bmc.2022.116946
51. Benassi A, Peñalver P, Pérez-Soto M *et al.* Structure-activity study on substituted, core-extended, and dyad naphthalene diimide G-quadruplex ligands leading to potent antitrypanosomal agents. *J Med Chem* 2024;67:10643–54. https://doi.org/10.1021/acs.jmedchem.4c00135
52. Pérez-Soto M, Ramos-Soriano J, Peñalver P *et al.* DNA G-quadruplexes in the genome of *Trypanosoma cruzi* as potential therapeutic targets for Chagas disease: dithienylethene ligands as effective antiparasitic agents. *Eur J Med Chem* 2024;276:116641. https://doi.org/10.1016/j.ejmech.2024.116641
53. Cantara A, Luo Y, Dobrovolná M *et al.* G-quadruplexes in helminth parasites. *Nucleic Acids Res* 2022;50:2719–35. https://doi.org/10.1093/nar/gkac129
54. Puig Lombardi E, Londoño-Vallejo A. A guide to computational methods for G-quadruplex prediction. *Nucleic Acids Res* 2020;48:1–15. https://doi.org/10.1093/nar/gkz1097
55. Bedrat A, Lacroix L, Mergny JL. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res* 2016;44:1746–59. https://doi.org/10.1093/nar/gkw006
56. Brázda V, Kolomazník J, Lýsek J *et al.* G4Hunter web application: a web server for G-quadruplex prediction. *Bioinformatics* 2019;35:3493–5. https://doi.org/10.1093/bioinformatics/btz087
57. Institute of Biophysics of the Czech Academy of Sciences. Bioinformatics tools server. http://bioinformatics.ibp.cz/. *Date accessed 1 April 2026*.
58. Hui WWI, Simeone A, Zyner KG *et al.* Single-cell mapping of DNA G-quadruplex structures in human cancer cells. *Sci Rep* 2021;11:23641. https://doi.org/10.1038/s41598-021-02943-3
59. Biffi G, Tannahill D, McCafferty J *et al.* Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature Chem* 2013;5:182–6. https://doi.org/10.1038/nchem.1548
60. Hänsel-Hertsch R, Spiegel J, Marsico G *et al.* Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nat Protoc* 2018;13:551–64. https://doi.org/10.1038/nprot.2017.150
61. Siegel TN, Hekstra DR, Kemp LE *et al.* Four histone variants mark the boundaries of polycistronic transcription units in

Downloaded from https://academic.oup.com/nar/article/54/12/gkag637/8721998 by University of York user on 17 July 2026

Trypanosoma brucei. *Genes Dev* 2009;23:1063–76. https://doi.org/10.1101/gad.1790409
62. Faria JRC, Tinti M, Marques CA *et al.* An allele-selective inter-chromosomal protein bridge supports monogenic antigen expression in the African trypanosome. *Nat Commun* 2023;14:8200. https://doi.org/10.1038/s41467-023-44043-y
63. Chen S, Zhou Y, Chen Y *et al.* fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34:i884–90. https://doi.org/10.1093/bioinformatics/bty560
64. Müller LSM, Cosentino RO, Förstner KU *et al.* Genome organization and DNA accessibility control antigenic variation in trypanosomes. *Nature* 2018;563:121–5.
65. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–9. https://doi.org/10.1038/nmeth.1923
66. Li H, Handsaker B, Wysoker A *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–9. https://doi.org/10.1093/bioinformatics/btp352
67. Feng J, Liu T, Qin B *et al.* Identifying ChIP-seq enrichment using MACS. *Nat Protoc* 2012;7:1728–40. https://doi.org/10.1038/nprot.2012.101
68. Ramírez F, Ryan DP, Grüning B *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 2016;44:W160–65.
69. Robinson JT, Thorvaldsdóttir H, Winckler W *et al.* Integrative genomics viewer. *Nat Biotechnol* 2011;29:24–6. https://doi.org/10.1038/nbt.1754
70. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26:841–2. https://doi.org/10.1093/bioinformatics/btq033
71. Lawrence M, Huber W, Pagès H *et al.* Software for computing and annotating genomic ranges. *PLoS Comput Biol* 2013;9:e1003118. https://doi.org/10.1371/journal.pcbi.1003118
72. Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. *Bioinformatics* 2021;37:1639–43. https://doi.org/10.1093/bioinformatics/btaa1016
73. Di Tommaso P, Chatzou M, Floden EW *et al.* Nextflow enables reproducible computational workflows. *Nat Biotechnol* 2017;35:316–9. https://doi.org/10.1038/nbt.3820
74. Rabuffo C, Schmidt MR, Yadav P *et al.* Inter-chromosomal transcription hubs shape the 3D genome architecture of African trypanosomes. *Nat Commun* 2024;15:10716. https://doi.org/10.1038/s41467-024-55285-9
75. Krzywinski M, Schein J, Birol I *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* 2009;19:1639–45. https://doi.org/10.1101/gr.092759.109
76. Rasche H, Hiltemann S. Galactic Circos: user-friendly Circos plots within the Galaxy platform. *Gigascience* 2020;9:giaa065. https://doi.org/10.1093/gigascience/giaa065
77. Kent J. (bigwigtowig. In GitHub repository. GitHub. 2021. https://github.com/ucscGenomeBrowser/kent. Date accessed 7 May 2026.
78. De Cian A, Delemos E, Mergny JL *et al.* Highly efficient G-quadruplex recognition by bisquinolinium compounds. *J Am Chem Soc* 2007;129:1856–7. https://doi.org/10.1021/ja067352b
79. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;30:923–30. https://doi.org/10.1093/bioinformatics/btt656
80. Chen Y, Chen L, Lun ATL *et al.* edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *Nucleic Acids Res* 2025;53:gkaf018. https://doi.org/10.1093/nar/gkaf018
81. Wickham H. *ggplot2: elegant graphics for data analysis*. New York: Springer-Verlag, 2016.
82. Powell D. Degust: interactive RNA-seq analysis. 2019. https://doi.org/10.5281/zenodo.3258932. Date accessed 18 October 2019.
83. Guhl F, Vallejo GA. *Trypanosoma (Herpetosoma) rangeli* Tejera, 1920: an updated review. *Mem Inst Oswaldo Cruz* 2003;98:435–42. https://doi.org/10.1590/S0074-02762003000400001
84. Hertz-Fowler C, Figueiredo LM, Quail MA *et al.* Telomeric expression sites are highly conserved in *Trypanosoma brucei*. *PLoS One* 2008;3:e3527. https://doi.org/10.1371/journal.pone.0003527
85. Maree JP, Tvardovskiy A, Ravnsborg T *et al.* *Trypanosoma brucei* histones are heavily modified with combinatorial post-translational modifications and mark Pol II transcription start regions with hyperacetylated H2A. *Nucleic Acids Res* 2022;50:9705–23. https://doi.org/10.1093/nar/gkac759
86. Xiao K, Zhang R, Tu J. Evaluating reproducibility and best practices for replicate design in G-quadruplex ChIP-Seq studies. *Int J Mol Sci* 2025;26:9769. https://doi.org/10.3390/ijms26199769
87. Wedel C, Förstner KU, Derr R *et al.* GT-rich promoters can drive RNA pol II transcription and deposition of H2A.Z in African trypanosomes. *EMBO J* 2017;36:2581–94. https://doi.org/10.15252/embj.201695323
88. Haddach M, Schwaebe MK, Michaux J *et al.* Discovery of CX-5461, the first direct and selective inhibitor of RNA polymerase I, for cancer therapeutics. *ACS Med Chem Lett* 2012;3:602–6. https://doi.org/10.1021/ml300110s
89. Bhartiya D, Chawla V, Ghosh S *et al.* Genome-wide regulatory dynamics of G-quadruplexes in human malaria parasite *Plasmodium falciparum*. *Genomics* 2016;108:224–31. https://doi.org/10.1016/j.ygeno.2016.10.004
90. Mao SQ, Ghanbarian AT, Spiegel J *et al.* DNA G-quadruplex structures mold the DNA methylome. *Nat Struct Mol Biol* 2018;25:951–7. https://doi.org/10.1038/s41594-018-0131-8
91. Zhang ZH, Qian SH, Wei D *et al.* *In vivo* dynamics and regulation of DNA G-quadruplex structures in mammals. *Cell Biosci* 2023;13:117. https://doi.org/10.1186/s13578-023-01074-8
92. Albanaz ATS, Gerasimov ES, Shaw JJ *et al.* Genome analysis of *Endotrypanum* and *Porcisia* spp., closest phylogenetic relatives of *Leishmania*, highlights the role of amastins in shaping pathogenicity. *Genes (Basel)* 2021;12:444. https://doi.org/10.3390/genes12030444
93. Spiegel J, Cuesta SM, Adhikari S *et al.* G-quadruplexes are transcription factor binding hubs in human chromatin. *Genome Biol* 2021;22:117. https://doi.org/10.1186/s13059-021-02324-z
94. Lago S, Nadai M, Cernilogar FM *et al.* Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nat Commun* 2021;12:3885. https://doi.org/10.1038/s41467-021-24198-2
95. Antariksa NF, Di Antonio M. The emerging roles of multimolecular G-quadruplexes in transcriptional regulation and chromatin organization. *Acc Chem Res* 2024;57:3397–406. https://doi.org/10.1021/acs.accounts.4c00574
96. Lima DA, Gonçalves LO, Reis-Cunha JL *et al.* Transcriptomic analysis of benznidazole-resistant and susceptible *Trypanosoma cruzi* populations. *Parasites Vectors* 2023;16:167. https://doi.org/10.1186/s13071-023-05775-4
97. Ashby EC, Havens JL, Rollosson LM *et al.* Chemical inhibition of bromodomain proteins in insect-stage African trypanosomes perturbs silencing of the variant surface glycoprotein repertoire and results in widespread changes in the transcriptome. *Microbiol Spectr* 2023;11:e0014723. https://doi.org/10.1128/spectrum.00147-23
98. Nuccio SP, Cadoni E, Nikoloudaki R *et al.* Chemically modified CRISPR–Cas9 enables targeting of individual G-quadruplex and i-motif structures, revealing ligand-dependent transcriptional perturbation. *Nat Commun* 2025;17:385. https://doi.org/10.1038/s41467-025-67074-z
99. Lin J, Hou JQ, Xiang HD *et al.* Stabilization of G-quadruplex DNA by C-5-methyl-cytosine in bcl-2 promoter: implications for epigenetic regulation. *Biochem Biophys Res Commun* 2013;433:368–73. https://doi.org/10.1016/j.bbrc.2012.12.040

Received: July 23, 2025. **Revised:** May 13, 2026. **Accepted:** May 16, 2026
© The Author(s) 2026. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Downloaded from https://academic.oup.com/nar/article/54/12/gkag637/8721998 by University of York user on 17 July 2026